The perception of intonation in native and non-native linguistic contexts and by different individuals: From question-answer categorization to the integration of prosody and discourse structure

Doctoral Course in Psychological Sciences and Education

XXX Cycle

Disciplinary Sector: L-LIN/01 Linguistics

Ph.D. Candidate

Alessandra Zappoli

Academic Year 2017/18

To all my family,
to my father,
to Luigi, Ada, Inge and Zeno.

Dearly, to
Diane and Jake

Acknowledgments

First and foremost, I would like to thank my family, for allowing me to study, for having supported me in all these years, during my journeys and years I have spent abroad from a young age, for never questioning my decisions, and for encouraging me during difficult times. In particular, I want to thank my father for the greatest support, and I want to thank and dedicate this work to the four little ones, who reminded me the joy and pleasure of learning, whenever I risked forgetting.

I believe language and foreign languages have been in my life from my childhood. I remember my first encounters with English, German and French during my school years, and I remember playing around with them, attempting to talk with "foreigners" on vacation and on the other occasions I had.
However, it is during the year I spent abroad as a foreign exchange student that I have made the decision to pursue the study of language and foreign languages. For this reason, I would like to thank the people I met in Duluth, in particular, my host-family Kim, Sam and Pat; and two special people that I will always remember dearly, Diane and Jake. In particular, I will always hold fond memories of my linguistic conversations in "Italish" with Jake, and his true talent as a second language learner.

Then, my interest and passion for languages bloomed in the years of my bachelor's and master's degrees at the University of Bologna. I want to thank Edoardo Vineis and Emiliano Guevara for inspiring me during the beginning of my formation, Mario Vayra and Cinzia Avesani for introducing me to, and forming me in the study of speech sciences.

I chose to focus on the perception and acquisitional mechanisms of language, turning my attention to the field of cognitive sciences and psycholinguistics during my Ph.D.
I want to thank, my advisor, Francesco Vespignani for believing in my project and giving me the opportunity to develop new competencies, covering many fields. Above all, I appreciated and gained from Francesco's multidisciplinary background, and I have learned a new perspective on my topics of interest, as well as new methodological skills during our thoughtful conversations. The quality of my reasoning undoubtedly increased, thanks to him, as well as my awareness on what means to be a scientist. Above all, I am extremely grateful for his moral support during these years.

For the help and support, I received during the developing of my project and the data collection I would like to thank Massimo Grassi, and again for her thoughtful advice Cinzia Avesani. Moreover, I would like to thank Vivian Grillo, who assisted me in finding the participants and running the experiments. She shared with me her enthusiasm and passion for the study of sound in language and music, and I have sincerely enjoyed it and learned from it.

A special thank goes to Petra B. Schumacher and her team for welcoming me at the University of Cologne. Her leadership and guidance helped me to refine my thinking and allowed me to develop several competencies that, as I came to believe, are essential in the formation of a scientist as much as the scientific ones. In particular, I would not have successfully completed my study during my visiting period without the support and collaboration of all the people I have met at the Institut für Linguistik and at the Institut für Linguistik – Phonetik.

Firstly, I would like to thank Martine Grice and Stefan Baumann for their advice, collaboration, and guidance. I would like to thank Silvia Dahmen for her time, her voice, and suggestions during the recording of the stimuli, and Francesco Cangemi for the smart tips during the preparation of the experimental material.
I would like to thank Lena Straßburger, Florian Bogner, Barbara Tomaszewicz-Özakın, Hanna Weiland-Breckle, Diana Dimitrova, Rebekka Wanka, Melanie Fuchs, Filiz Oezden and Claudia Kilter for their help and support, their time, for having shared their competencies and know-how with me, and for the good time in Cologne.

My sincerest thank goes to all the people with whom I shared my Ph.D. experience, joy, pain, and panic. Thanks to my office-mate Marco Bressan, to whom I owe the developing of my statistical skills, and without whom, the Ph.D. program would not have been the same. Thanks to Diego Azevedo Leite with whom I shared great moments exploring Trentino and engaging in deep philosophical conversations. Thanks to Nicoletta Biondo for the great discussions about linguistics and good times in Rovereto; and to Giulia Calignano for her true passion in science, and the fruitful exchange of opinions, points of view, and knowledge in the cognitive sciences and in the study of language acquisition. Thanks to Andrea Beretta and Michele Giannotti for the nights out rescuing the world, and the developing of a common scientific ground; and to Giuditta V. Smith for her kind support and sharing of new ideas debating language topics.

I would also want to thank my dearest and closest friends that I have met along the years, during school, university and during the periods I have spent abroad. I want to thank my dearest friend Elisa, with whom I have shared so many experiences since we were teenagers, who has always been there with a smile and a deep thought whenever I needed it. For the same reasons, and a friendship lasting as well over 20 years, a special thank goes to Michel and his peaky points of view that can be lovely annoying, but never trivial.
I am incredibly thankful for the support I have received during my university and Ph.D. years from Sara, with whom I shared the passion for foreign languages and the panic for exams, and funny episodes during our trips, and with whom I have always exchanged thoughtful opinions.
A huge thank you to Claudia, whom I have randomly met in Heidelberg in one of the best years of my life, finding out we came from the same good old town, with whom I shared the craziest stories about the craziest people, and with whom I truly feel a citizen of the world. Distance does not matter.

I want to thank all the people I have met in Würzburg, in another curious set of circumstances. It was the year right before I started the Ph.D. program, and I was young, inexperienced and hungry, and for some reasons I did meet people who contributed making that time, one of the most scientifically stimulating and creative of my life. In particular, I want to thank Maria Luisa Lamberto, for the advice she gave me in teaching Italian as L2, her vast and deep knowledge in philosophy and literature, and mostly for teaching me resilience while reaching our goals. Finally, I express my gratitude to Chiara Esposito, who showed me a part of my country I did not know before, discussed mathematics with me as if I were a mathematician, and indicated me the path as a good advisor, in a time in which I was, in truth, clueless about the academic life.

In conclusion, I want to thank Oliver Niebuhr and again Petra B. Schumacher for their comments and indications on a first version of the manuscript. All remaining errors are, of course, only mine.
Last but not least, I want to thank all the participants to the studies, for their time, patience, and effort. Truly, without them, none of this would have been possible.

Table of Contents

List of Figures

List of Tables

Introduction and outline of the thesis

Pitch in language expresses both linguistic and paralinguistic functions, and it operates both at lexical and at sentence-level. At the lexical level, in tonal languages, pitch has a lexical function: it acts as a distinctive feature carrying information that is necessary to disambiguate homophonic lexical entries, and it allows to access one specific meaning. At the sentence-level, in intonational languages, pitch is one of the suprasegmental features that combine into a complex structure, allowing certain portions of speech to stand out within discourse and to change sentence modality. At the sentence-level, pitch also has a paralinguistic function, called "emotional prosody": the pitch range, together with intensity, the use of pauses and durations conveys to the listener the emotional state of the speaker. The combination of pragmatic and paralinguistic information allows disambiguating whether the speaker is serious, happy, or angry, and disentangles if the utterance has to be interpreted literally, sarcastically, or as a joke.

Pitch is defined as the perceptual correlate of the Fundamental Frequency (F0) within a complex sound wave. It is the perceptual sensation of a rising tone associated to the increase of F0, and the sensation of a lowering tone associated to the decrease of F0 (Raphael, Borden, & Harris, 1994). Apart from serving a linguistic function, it is primarily a perceptual dimension also belonging to the domain of music and singing.

By observing the several pitch function and interpretations, it is fascinating to see how broad its area of application in communication is, and how large is the conceptual distance between each of these functions. At an abstract level, two F0 modulations could be processed as two different musical melodies, as two Chinese tones allowing to access either the meaning of the word “jade” (pinyin transcription: yù – 4th tone) or of the word “fish” (pinyin transcription: yú – 2nd tone), as a statement opposed to a question, as an emotional state having lost patience or being doubtful about a situation.

By focusing on the linguistic functions: what allows to switch from the auditory perception of a rising frequency, to the recognition of the word “fish” in Chinese, or to the decoding of the pragmatic instruction, conveying that the speaker terminated the conversational turn and waits for a reply – recognizing a question? In intonational languages, is the same rising movement undoubtedly a question for every speaker of a language? And if yes, what undoubtedly makes it a question? And across languages, when speakers acquire a second language (L2) later in life, is the pragmatic function of an F0 rise undoubtedly a question, if, for example, in the native language of the speaker, questions are uttered with a descending contour?

Do speakers of different languages process an F0 rise and decode its pragmatic function in the same way?

Pitch and intonation are a part of the phonological system that a learner needs to acquire when approaching an L2. The phonological aspects of the L2 are the aspects of Second Language Acquisition (SLA) that are more difficult to learn by late/adult learners. Moreover, their phonetic implementation in the L2 speaker's speech frequently allows a native speaker to recognize whether an L2 learner has a foreign accent. The deterministic linguistic approach assumes that, despite several years of training or exposure, almost no L2 late learner can ever reach a native-like mastering of the L2 phonologic segmental and suprasegmental features.

More recent developments of research on language aptitude began to investigate cases in which, despite all the odds, late L2 learners indeed show a native-like L2 pronunciation, including suprasegmental aspects. From the evidence in production, it became relevant to identify which are the learning mechanisms -

starting from perception - that allow these talented individuals to maintain the ability to acquire the sound properties of an L2 in adulthood.

Since the processing of intonation entails the ability: to appropriately integrate the acoustic features with word-meaning and with the syntactic structure of the utterance, and to deal appropriately with the exchange of information between interlocutors, a specific talent in the processing of intonation can be a signal of above-average skills in the processing of pragmatic aspects of language. A talent in processing L2 suprasegmental features in late learners can be a predictor of above-average abilities in pragmatic processing. Adopting a specular perspective, a particular deficit in the decoding of pragmatic aspects in language can be a predictor of non-efficient processing of intonation and construction of the discourse model. Aside from the possibilities of causal links between the two competencies, a correlation between pragmatic competence and talent in (L2) intonation processing can thus be hypothesized.

A population of interest, reported to show deficits in the decoding of pragmatic aspects of language, and that is often not included in the linguistic investigation, is the one affected by atypical development and diagnosed with Autism Spectrum Conditions (ASC). ASC are neurodevelopmental conditions that are characterized by social and communication difficulties, and by the presence of unusually narrow interests and stereotyped patterns of behavior, that include the resistance to change (APA, 1994; Ruta, Mazzone, Mazzone, Wheelwright, & Baron-Cohen, 2012). The variability with which cognitive and linguistic abilities vary in the population led to the notion of spectrum disorder. By adopting the notion of the spectrum, the population distributes ranging from people displaying normo-typical social and cognitive abilities, to individuals displaying atypical behavior. The disorder can be circumscribed to the social sphere, or in the extreme cases, it can also imply the presence of severe mental retardation (Rice, Warren, & Betz, 2005; Volkmar, Lord, Bailey, Schultz, & Klin, 2004).

These constructs, developed in the field of pathological conditions, are indeed a source of information for linguistic research. For example, the analysis of the deviating patterns in intonation is actually used as a diagnostic criterion in the clinical assessment of ASC (Diehl, Watson, Bennetto, McDonough, & Gunlogson, 2009; Le Couteur, Haden, Hammal, & McConachie, 2008; Lord et al., 2000; Rutter, Le Couteur, & Lord, 2003). Conversely, one can hypothesize that ASC traits or measures in the neurotypical population can be used as a possible predictor of talent or deficits in dealing with prosodic aspects of language that have essential roles in pragmatics.

These two groups of individuals, one showing exceptional talent in processing pitch and pragmatic functions, and on the other demonstrating an impairment, can be considered as two opposite tails of a single distribution, representing the ability to appropriately master the voice modulation to express linguistic and paralinguistic information in speech.

The phonological and phonetic description of intonation across languages, its perceptual processing, and organization, its function in the construction of a discourse model, and its implementation in a second language are often separately and independently addressed, sometimes by different scientific communities (linguists, psychologists, neuroscientists). Each literature is often biased on its own different core concepts and internal theoretical debates, and thus, a comprehensive view of the phenomenon is hard to be captured within each separate disciplinary field.

The main aim of this thesis is to investigate, in normo-typical adults, the relevant factors, allowing or impairing the correct pragmatic interpretation of pitch movement in discourse, in native and second language, and to consider the results in a broad perspective that overcomes strict monodisciplinary borders.

This will be done in 3 experiments. Two psychophysical behavioral experiments are devoted to the study of pitch processing at a sensory level a third experiment will study how congruent and violating pitch movements are integrated within discourse, at a higher level, in L1 and L2, through the neurophysiological technique of Electroencephalography (EEG).

All 3 experiments manipulate pitch in relation to its pragmatic functions (question vs. statements disambiguation and information packaging at the discourse level), and all experiments manipulate pitch in function of the access to the meaning of the target elements to which it is aligned. How a non-native language impacts on pitch processing is central in Experiment 3, but it is present also in Experiment 1 as an intermediate experimental condition.

In all 3 experiments, the cognitive aspects at the individual level and related to the presence of an efficient or inefficient processing of the pragmatic functions linked to pitch modulation are captured by the adopted questionnaires: the Autism Spectrum Quotient (AQ) (Baron-Cohen et al., 2001) and the Interpersonal Reactivity Index (IRI) (Davis, 1983, Albiero et al.,2006). These two questionnaires are composed overall of 9 subscales (for a more extensive description of the scales see Paragraph 1.6.3.). These two scales have been chosen because they can grant a quantitative numerical measure of the individual traits of the participants considered relevant for the processing of the pragmatic function of speech and they provide a reference to what is recognized as atypical processing, belonging to the Autism Spectrum Conditions (ASC). In particular, two specific subscales of the AQ indicate the presence of atypical functioning of the attentional mechanisms. These measures will be used to study the presence of individual variability correlated with the efficient processing of pitch movements and their function in discourse.

One specific aim of the research is to try to adopt a multidisciplinary perspective, by giving similar weights to the different theoretical and experimental aspects involved in this research. This requires to consider the merits and limitations of assumptions coming from each of the presented areas and the implications they have on one another. On one side, the processing of intonation involves several levels of linguistic investigation, from phonetics to pragmatics, on the other side, pragmatics from a psychological perspective is observed and measured in terms of appropriate, typical and efficient social interactions between individuals. Pragmatics: is the ability to decode the interlocutors' mental states (Theory of Mind - ToM), the ability to imagine and empathize with others' conditions, the ability and to successfully and comfortably conduct a conversation. Impairments in the area of pragmatics, on the other hand: are related to the difficulty of decoding implicit and inferable information; are related to the higher allocation of attention on the details of the physical signal and difficulties in the understanding of the context of use; and are related to the feeling of distress in social situations.

For this reason, the empirical part of the work will be preceded by an introductory chapter in which the theoretical background of concepts that are relevant for the three experiments are discussed critically, also by briefly highlighting the historical development of the fields and their interplay.

The thesis is thus structured in the following way:

Chapter 1 introduces:

- the Autosegmental Metric (AM) theory of intonation, which is adopted as the general theoretical framework, as well as the phonological inventories (ToBI systems) of the two languages under investigation (Italian and German);
- the notion of category and categorical perception and the evidence of its application in the field of intonation;
- the main concepts in the field of SLA, and how perception is known to vary in L2 when compared to native-language processing;
- the main evidence about individual variability and the adopted questionnaires.

Chapter 2 reports the first psychophysical experiment that specifically addresses the topic of Categorical Perception of Boundary Tones (BTs) in Italian. The presence of CP in intonation is controversial since there is evidence that at the same time supports it and dismisses it. The currently more supported position is that intonation is quasi-categorically perceived. I suggest that a relevant factor modulating perception from continuous to categorical is played by segmental and lexical aspects co-occurring with pitch and that, at a later stage, are integrated with pitch. The presence of categorical perception can be necessarily linked to the presence of specific segmental and lexical information, and the efficiency with which segmental and suprasegmental features are processed and integrated can vary across individuals – some decoding, more efficiently than others, the pragmatic function linked to the pitch movement. In this experiment the amount of information carried on by the segmental material is modulated in 4 levels, ranging from high to low recognizability. The experiment is performed in two experimental orders allowing to observe two perceptual strategies that normally are not observable in a CP paradigm, and that may be in my view relevant to disentangle whether pitch is linguistically processed or not.

Chapter 3 reports, by adopting only two experimental conditions of the previous experiment (highest and lowest linguistic recognizability of segments), the application of the full Categorical Perception (CP) paradigm to the perception of BTs. Concerning the original paradigm, I also introduced a different discrimination procedure that estimates the minimum perceptual threshold of pitch movements within and across prosodic categories with the aim to have a more fine-grained measure of individual sensitivity to small changes in tones.

Chapter 4 presents a language comprehension experiment conducted in German, with native speakers and second language adult learners, native speakers of Italian. The experiment is conducted adopting the EEG technique and investigates the online integration of Nuclear Pitch Accents (NPA) with the underlying discourse structure, explicitly focusing on the resolution of anaphoric relations. The aim in this experiment is to establish how and when the processing in L2 speakers diverges from the processing of native language speakers and evaluate whether the differences affect the phonological decoding and later stages of discourse information structure integration.

Chapter 1
An intonational big picture

1.1. A Phonological model to describe intonation

1.1.1. Autosegmental Metric theory of Intonation (AM) and the ToBI system

Pitch, as perceptual correlate of the Fundamental Frequency (F0), is a cue for both linguistic and paralinguistic information in speech. Its organization in each language gives rise, together with other phonetic parameters (e.g., intensity, duration, spectral information), to what it is referred to as the intonational system. Intonational phenomena belong to the suprasegmental features in speech. They, thus, set the constraints of the domain of interest at the sentence level.

In this thesis, intonation is described adopting the Autosegmental Metric theory of intonation (AM)(Bruce, 1977; Ladd, 1996, 2008; Liberman & Prince, 1977; Pierrehumbert, 1980). The AM approach allows analyzing the continuous variation of the F0 signal in discrete phonological representations. Within AM, intonation is a system always composed of two levels: the segmental and suprasegmental one, and by the rules that associate these two levels. F0 can be decomposed in phonological units only considering its association to the segmental level, which is metrically and hierarchically organized in prosodic units.

The F0 contour within the Pierrehumbert model is composed by a sequence of peaks and valleys corresponding to high (H) and low (L) level tones. Ramps and falls are labeled adopting a combination of the original level tones, creating bitonal units: from low to high (L + H) and from high to low (H+L).

The hierarchical organization of the segmental level has the syllable as its minimal unit, mandatorily composed of at least one vowel, and then it spans to the highest level comprehending the whole utterance, through the intermediate prosodic units: the foot, the phonological word, the intermediate phrase (ip) and the intonational phrase (IP)(Beckman & Pierrehumbert, 1986; Ladd, 1986). The original model of Pierrehumbert was strongly relying on the notion of foot which is the distance between one stressed syllable and the following one(Hayes, 1980; Selkirk, 1980). In the course of the years, different models of hierarchical structures have been proposed introducing further intermediate prosodic units [1]. The phonological word is a domain in which the intonational events can spread within the same lexeme across morphemes, including clitics (Hall, 1999). The intermediate phrase (ip), officially introduced by Beckman and Pierrehumbert (1986), refers to a prosodic unit larger than a phonological word but still dominated by the intonational phrase (IP). There are cases in which, an IP can be further subdivided into smaller units, but these smaller units are not delimited by long pauses and breaks, that usually delimit constituents at the higher IP level. Finally, the highest level of the hierarchy is represented by the whole utterance, that can correspond to an IP if uttered in one single prosodic unit.

The Intonational Phrase (IP) represents the typical domain of investigation of intonation. The criteria required for the identification of an IP are summarized by von Heusinger (1999, p. 72):

1) Timing: a pause can precede and follow an IP;
2) Metrical: marking, e.g., the position of the metrical head;

[1] The definition of "intermediate phrase" is based on tonal properties only within Beckman & Pierrehumbert model, and also considers rhythmic aspects, such as the presence of a pause, in Ladd's description. Since the theoretical implications of the two models are not relevant in this work, their differences are not further discussed. A detailed model of the prosodic hierarchy is also proposed by (Nespor & Vogel, 1986; Selkirk, 1986)

3) Tonal: the presence of BTs.
4) Junctural: the boundaries of an IP can block junctural phenomena across prosodic units.
5) Syntactic-prosodic: the boundaries of an IP are mapped with the syntactic structure.
6) Semantic: the content of an IP must constitute a unit of sense.

The syllables within a prosodic unit do not have equal weight; on the contrary, within each multiword-unit, a single syllable always stressed at the word level, is metrically stronger becoming the metrical head of the prosodic unit.

In the original model, this relation is proposed as an alternation between strong (s) and weak (w) syllables across feet. The metrical head, thus, is the syllable marked as strong which is not dominated in the hierarchical structure by any other weak syllable (see an example adapted from Pierrehumbert, 1980, p.33 in Figure 1.1.1).

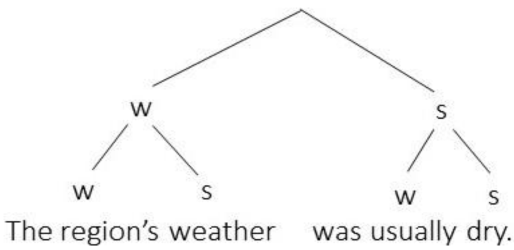


Figure 1.1.1 Alternation of strong and weak syllables

The resulting metrical structure of the IP in the example can be then finally formalized as in Figure 1.1.2:

```
                                    *
             *                      *
[ [σ  σ  σ  σ  σ]ip  [σ  σ  σ  σ  σ]ip ]IP
             T*  T-                 T* T- T%
```

Figure 1.1.2 Metrical structure of an IP

Languages can differ in their positioning of the head of the prosodic unit. Some languages have the metrical-head on the right or on the left. In right-dominant languages, the head corresponds to the last stressed syllable on the right of the prosodic unit; vice-versa in left-dominant languages, the metrical head is positioned on the first stressed syllable of the prosodic unit.

The tune-text association is granted by the *alignment* of the F0 contour with the metrically organized segmental level generating a linear tonal structure of tonal events. Tones aligned to stressed syllables assume the status of Pitch Accents (PAs), PAs aligning to the metrical head represents the most prominent elements within the prosodic unit and are referred as *Nuclear PAs* (NPAs). Tones aligning to the edge of an intermediate prosodic unit (e.g., the intermediate phrase) are referred as Phrase Accents (PhAs), and tones aligning to the edges of the prosodic unit assume the status of *Boundary Tones* (BTs). A tune is therefore composed of at least one pitch accent, (in the case this is the only accent within a unit it can also assume the status of NPA), a phrase accent and a boundary tone.

Syllables that are pitch-accented acquire acoustical salience and, depending on the referring language system, they can be characterized by a longer duration. The perceived prominence of a specific syllable, within the AM system, therefore, is the combination of two separated independent tiers: pitch and stress.

Overall the structure proposed by the AM theory is composed of 3 tiers: the segmental one, in which at least one vowel needs to serve as an anchor point; the tonal tier, where for each vowel there has to be at

least one tone; and the association tier that contains the mapping rules of tones and segments (Goldsmith, J., 1976).
The tune-text association is signaled in the transcription by specific labels.
The alignment of a tonal movement to a stressed syllable is marked with a * (star), the alignment with the edge of an intermediate prosodic unit with a dash (-), and the alignment with the edge of the utterance is marked with the symbol % as in figure 1.1.3 below, adapted from Pierrehumbert (1980, p. 29):

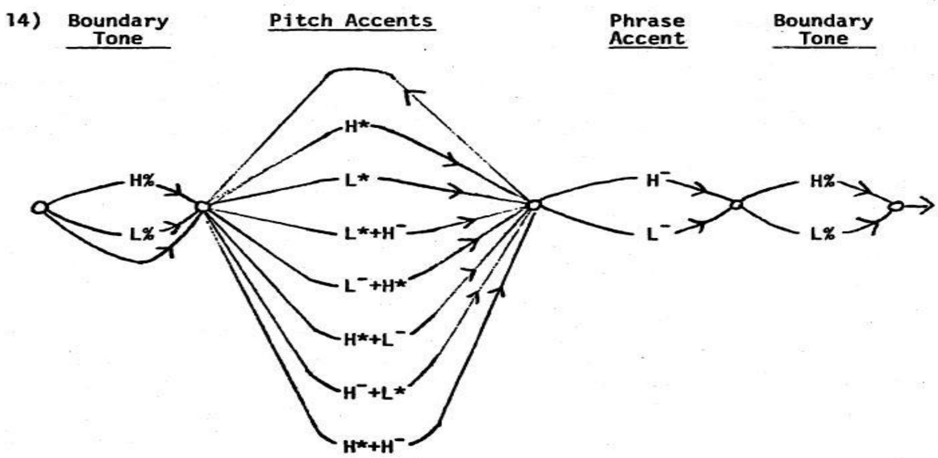


Figure 1.1.3 Pierrehumbert's tune-text association system

In Figure 1.1.3, it is reported the original phonological inventory of PAs, PhAs, and BTs proposed by Pierrehumbert (1980) for English. The labels H* and L* represent level tones, while the labels H*+L, H+L*, L*+H, L+H* and H*+H represent bitonal units. In addition, two phrase accents are present H- and L-, and two BTs: H% and L% signaling that utterance begins or end respectively with high or low tone.
The transcription of the intonational contour adopting AM has been then formalized with the Tone and Break Indexes (ToBI) system in the early 1990s. The ToBI system is conceived as a transcription containing several tiers of analysis, in which it is possible to annotate the speech's orthographic transcription and the suprasegmental events aligning with it. It combines the aspects strictly pertaining to the annotation of the intonational contour and the aspects pertaining to the prosodic organization of the segments, such as the marking of pauses and of the hierarchical structure. The ToBI system allows for a standardization of the inventory of PAs, PhAs, and BTs. The first ToBI system was specifically introduced to describe the phonological inventory of American English: the MAE ToBI (Mainstream American English ToBI). In this formalization, the original inventory proposed by Pierrehumbert is reduced to 5 PA types with the introduction of a diacritic symbol (!) marking the presence of downstep (Ladd, 2008). The concept of downstep was already present in the original Pierrehumbert system and accounts for a sequence of two H tones, where the second target shows to be lower than the first one but not labelable as L. The notion of downstep allows to account for declination effects along the utterance and local lowering phenomena, overcoming the limitations of strict binary classification. Only high tones, therefore, can be downstepped. The 5 PAs in the MAE ToBI are H*, L*, L+H*, L*+H, and H+!H*. The description of phrasing uses the original analogous marking of phrase accents and boundary tones proposed by Pierrehumbert (1990).
The hierarchical organization of prosodic constituents is then granted by the Break Index (BI) annotation that labels each orthographic words' disjunction from 0 to 4.
Break Indices (Beckman & Elam, 1993; version 3.0 updated in 1997; p.31):

> "represent a rating for the degree of juncture perceived between each pair of words and between the final word and the silence at the end of the utterance. They are to be marked after all words

that have been transcribed in the orthographic tier. All junctures — including those after fragments and filled pauses — must be assigned an explicit break index value; there is no default juncture type."

Thus, the indices indicate the perceived separation between prosodic constituents from the word to the IP level. The indices that annotate the presence respectively of an ip and an IP are the ones labeled with the highest numbers (3) and (4). The lowest index (0) is rarely used, marking phenomena belonging to some specific types of speech, such as the broadcasting one, to indicate the presence of connected speech. The index (1) is the lowest level most commonly used, representing the "default" marking of word boundaries in phrase-medial position; and the index (2) is used to mark the presence of an incongruency between the presence of a perceived pause and the absence of a tonal event, which should typically be present.

1.1.2. The annotation of German intonation and the adoption of GToBI

Following the introduction of the AM approach and of the MAE ToBI system, different annotation systems have also been proposed for the transcription of German intonation. In the '90s both (Wunderlich, 1988) and (Uhmann, 1991) proposed their models. Wunderlich's model, despite adopting the main concept of AM (the distinction between a phonological abstract representation and a phonetic implementation) mainly focused on a phonological description. He proposed 6 accents: H*, H*HL* (bridge accent), %HL* (resembling the later introduced H+L*), L*H%, L* H (H%), H* H. The peculiar aspect of his system was the presence of 3 types of accent - PAs, BTs and non-BTs - and the formalization of phonological and binary oppositions pertaining the BTs. In his system, a BT could or could not be present, and could be either high or low. He also introduced the concept of markedness defining the BTs and non-BTs are low when unmarked and high when marked, and PAs are low when marked and high when unmarked.
Uhmann (1991) proposed a description with 4 PAs: H*+L, H*, L*+H, and L* and two BTs: H% and L%. BTs were assumed to have two levels also in offset of an IP, and the ip level was not represented. Since both Wunderlich and Uhmann models represented essential steps in the developing of a common standardized system of German intonation but are not the currently most adopted systems, the systems will not be further described here, but for a more detailed review, see (Grabe, 1998).
A comprehensive annotation system was followingly proposed by (Féry, 1993). In this work, intonation was also put into relation with pragmatic aspects of discourse, and information structure. In her system, she declares 3 NPAs: H* L, L* H, and L*HL, and no phrase accent is needed. Féry, (1993) proposed that pitch movements between the NPA and the BT could be accounted as trailing tones and were not necessary to describe the ip. BTs were present only in final position but not at the beginning of the IP, and a low BT was not necessary to be marked since the annotation of falling contours is already sufficient. (Féry, 1993)'s model also agreed with the studies of Kohler (1987, 1991) on Categorical Perception of alignment properties in intonation. In particular, she supported the evidence that the difference between an early and middle peak in a falling contour must be categorically perceived, while the difference between a middle and late contour must not. Kohler,(1987, 1991), in fact, introduced the notion of *early peak* for high modulations in pre-accentual position.
In 1995, the ToBI system was adapted to be suitable for the transcription of German intonation (Grice, Reyelt, Benzmuller, Mayer, & Batliner, 1996) with the name of GToBI.
The ToBI system is composed by at least 3 *tiers*, that comprehend labels for segmental information, tones, and break indices. In the first tier, it is usually reported the transcription of the uttered speech, on the second one there are the labels for PAs and Boundary Tones, and in the third one, the boundaries' strength

information (originally taken from MAE ToBI system) which corresponds to intermediate and intonational phrase boundaries (3 e 4). The phonological inventory comprehends the 5 PAs formalized within the MAE ToBI: H*, L*, L+H*, L*+H, and H+!H* with the addition of the falling H+L* PA. Downstepping is also present with the analogous symbol '!' and it is also possible to mark upset with the symbol ^. Differently, from the ToBI developed for English, GToBI does not mandatorily require a PA within an ip, and contrary to (Féry, 1993) it marks phrase accents as either low or high L- and H-, and contrary to Grabe (1998) marks BTs as either low or high. PhAs and BTs can combine in final position to express final rises and final falls.
The system accounts for leading tones displaced before the accented syllable. A high tone preceding the accented syllable is considered an early peak and a valley forms with the following high tone a rising onglide. This description results in two possible annotations of early peaks: H+!H* and H+L*. The first one is the early peak described by Kohler and transcribed as H+H*+L by (Féry, 1993) and the second one corresponds to the contour that Grabe (1998) analyzes in her dissertation focusing on the effects of downstep.
The rising onglide L+H* is described as a pattern that jumps up (Fox, 1984; Grice, Baumann, & Benzmüller, 2006) to the high-level tone on the vowel of the accented syllable and that is perceived as higher than the pitch on preceding syllables.
This accent, nonetheless, is at the center of some controversies since its early introduction. In fact, the peculiar aspect of the ToBI system is that the transcriber operates a linguistic decision when labeling a contour and identifies from phonetic features the phonological category present in the inventory. The difference between H* and L+H*, as well as the effect of downstep and position of the peak within the syllable, as Grabe (1998) highlighted, can be subtle since it mainly relies on the phonetic realization and alignment of the leading tone, as schematically represented in the table below.

The following table reports the combination of PA, PhAs and BTs adapted from Grice et al. (2006).

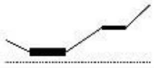
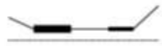
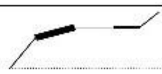
Description	GToBi	Contour	Context
Fall	H* L- %		Neutral statement Neutral W-question
	L+H* L-%		Contrastive assertion
Rise-Fall (late peak)	L*+H L-%		Self – evident assertion Sarcastic assertion
Rise	L* H-^H%		Neutral Y/N question Echo question
	L* L-H%		Indignation Answering phone
	L+H* H-^H%		Follow-up question
Level	L+H* H-(%)		Incompleteness Ritual expression
Fall-Rise	H* L-H%		Polite offer
Early Peak	H+!H* L-%		Established fact
	H+L* L-%		Soothing/Polite request
Stylised Step Down	(L+) H* !H-%		Calling contour

Table 1.1.1 GToBI System

Grice et al. (1996) investigated the Inter-transcriber consistency in PA labeling in GToBI, finding a percentage of 71%. One-third of the disagreeing judgments pertained the labeling of L+H* and H*. In addition, the two accents happened to be merged in one representation in studies in English adopting the MAE ToBI.

1.1.3. The annotation of Italian intonation and the adoption of ToBIt

Italian intonation is part of a bigger and extremely complex picture which deals with the description of the Italian language's phonological representations and phonetic realizations. Italian spoken language has been subjected to the process of language evolution responsible of the formation of all Romance languages from a common Latin root and has been influenced during its evolution process from other languages (given the historical, cultural, and political situation of the Italian territories in those years). The way the spoken language evolved did not similarly and homogeneous affected the spoken language within the whole country. This evolution mechanism result is the presence on the territory of different linguistic systems, which qualify as dialects and some in cases as recognized minority languages, that formed both a substratum for the Florentine variety, from which the standard written national language developed, and parallelly nowadays coexists with it.

The first studies on Italian intonation relied on the variety spoken in the region of Tuscany (Lepschy & Lepschy, 2013) and on broadcasting speech assumed as standard Italian (Ames, 1969; Avesani, 1990). In

recent years, then, several studies annotated the intonational systems in different varieties (for a fully detailed review of the analyzed varieties see (Barbara Gili Fivela et al., 2015; Grice, D'Imperio, Savino, & Avesani, 2005; Savino, 2009, 2012) with the aim to develop a common inventory of PAs and BTs to allow for both: cross-variety comparison and generalization. (Janet Pierrehumbert & Hirschberg, 1990)
The Italian version of the ToBI system is labeled ToBIt and has been formalized by Avesani (1995). The original system is provided with 3 NPAs: two monotonal, H*, L*, and three bitonal, H+L*, L+H*, L*+H; and with the following combination of intermediate phrase accents and BTs: L- L%, H- H%, L- H%, H-L%.
The BT, in sentence-initial position, is described adopting only one tone (H%), and it is proposed as relevant phonological marking of exclamative utterances and opposed to initial non-marked initial pitch movements uttered within the non-marked pitch range of the utterance.
Recently, in Gili Fivela et al. (2015), a comprehensive review of the attested phonological inventory of PAs and edge tones across variety has been collected. I adapt the description presented by Gili Fivela and coauthors in the table below.

The diacritics: ‘<’ refers to late peaks, ‘>’ refers to early peaks, and ‘¡’ refers to extremely high peaks.
Since in this work I restrict the participants in two experiments to speakers coming from an area positioned at the north-east of the Garda lake in the north of Italy, and in one experiment investigating L2 acquisition I select participants ranging from north to southern areas, the specific phonological inventory of each variety and its phonetic implementation is not of specific interest. Nonetheless, the overall attested inventory of spoken Italian is relevant in order to set specific predictions in the CP experiments and for the cross-linguistic comparison involving perception in native and second language. Table 1.1.2 reports the phonological inventory of PAs in Italian across varieties (adapted from Gili Fivela et al. 2015)

Description	ToBIt - PAs	Contour	Possible Contexts across varieties
Level	L*		Postfocal Prefocal
	H*		Y/N questions Postfocal Orders Exclamatives
Fall – Early Peak	H+L*		Broad focus Final items in lists Disjunctive question Wh-questions Orders Exclamatives
Rise-Fall	H* + L		Contrastive-corrective focus Y/N questions
Rise-Peak	L+H*		Contrastive-corrective focus
Late Peak	L+<H*		Vocatives Prenuclear accent
	L+¡H*		Incredulity Counterexpectational wh-questions
	L*+H		Y/N questions Wh-questions
	L*+>H		Exclamatives

Table 1.1.2 PAs in ToBIt

In the following Table 1.1.3, are reported the attested edge tones across varieties (adapted from Gili Fivela et al., 2015).

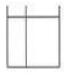
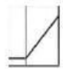
Description	ToBIt – Edge Tones	Contour	Possible Contexts across varieties
Level	L- or L%		Statements Questions
	H- or H%		Statements Lists Continuation Wh-questions
	H!H%		Vocative
Rise	LH%		Continuation Y/N questions Wh-questions
	L!H%		Countexpectetational Y/N questions
Fall	HL- or HL%		Y/N questions Orders Contrastive-focus

Table 1.1.3 Edge Tones in ToBIt

1.2. How intonation maps with discourse information structure

In the following section, I describe how pitch modulation has been linked to specific functions in speech, before and after the adoption of the AM modeling. It is possible to address this topic focusing on two main perspectives: i) the fact that intonation is present in speech independently from the language of use; thus, it must have a universal function in oral communication, linked to biological features of the vocal tract; ii) the fact that languages can show typological differences that can be expressed at different levels (see Ladd, 1996; and Mennen 2015 in section 1.5.3.)

Before addressing the question of how intonational categories are processed and acquired in L2, it is relevant to address: i) which aspects of the mapping of pitch modulation with discourse information are language-dependent; ii) which aspects are shared by individual independently from typological differences and iii) what different theoretical models imply. I will in particular address the role of AM theory, that specifically positions its basic unit, the tone, and its combination in Pitch Accents, within the phonological domain, and I will address in detail the literature concerning the association between pitch modulation and discourse information structure in West-Germanic and Romance languages, since the focus will be in the next chapters on German and Italian.

1.2.1. Universal vs. linguo-specific function of intonational movements.

Before the development of the AM approach, the intonational contour was either considered a whole, inseparable unit - holistic approach (Bolinger, 1989) – or decomposable in smaller units, that, nonetheless, had only a phonemic value – e.g., British School (Crystal, 1969; Halliday, 1967). Within these two perspectives the meaning of an intonational contour was not the combination of smaller meaningful units (Bolinger, 1964; M. Y. Liberman, 1975), but rather a holistic concept expressing, for example, modality (M. Liberman & Sag, 1974; Sag & Liberman, 1975), and attitude and emotion (O'Connor & Arnold, 1973). These interpretations were based only on the assumption that intonation reflected a universalistic function guided by two mechanisms (Ladd, 1996): a pure biomechanical one (A. M. Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967) and the assumption of the existence of a universal sound symbolism (Bolinger, 1978, 1989; Cruttenden, 1981).

According to the biomechanical perspective, the universal mechanism guiding the organization on intonation in prosodic constituents is constrained by the necessity of the speakers to breathe while producing speech. There appears to be a natural correspondence between prosodic units, representing a "breath group" and the conveying of meaning. An IP is an entity. The act of phonation itself reflects its dependence from breathing since the subglottal pressure is higher at the onset of a breath group, causing F0 to decline naturally – declination effect – along with an IP. In this sense, a decreasing pitch contour signals the non-marked end of a speech act and conveys finality. On the contrary, a final rise can be interpreted as a voluntary intention of the speaker to contrast the non-marked declining pattern, and therefore, it can be pragmatically interpreted as either: the intention to continue the speech, or the intention to pass the conversational turn to the interlocutor, as it happens in questions. In the same way, a local rise, or rise-fall within a prosodic unit, would represent a voluntary deviation from the declining natural pattern, and this, therefore, can be interpreted as a signal of relevance.

According to the second sound-symbolism perspective, apart from biomechanical reasons, high and low pitches oppose on the basis of an assumed prelinguistic meaning (Ladd, 1996). This approach comes at the end to interpretations that are similar to the ones based on the biomechanical approach, but the underlying reasons are different. In this perspective, high pitch signals incompleteness, tension and interest, and it is opposed to low pitch signaling, again, finality and completeness. In some cases, the difference in the explanation provided by the two approaches affects the interpretation assigned to specific tonal events. For example, a high Boundary Tone aligned in onset of an IP is explained referring to the higher subglottal pressure, adopting a biomechanical perspective, or referring to a prelinguistic intention to generate interest in the interlocutors, in the symbolistic one. Bolinger (in Ladd, 1996; p.114) defines 3 fundamental principles that according to his theory are common across languages:

1. The tendency of pitch to drop at the end of an utterance, and to rise at major breaks where the utterance remains incomplete;
2. The use of higher pitch in questions. Since in questions, the speaker expresses interest, and since the exchange is incomplete until the addressee answers;
3. The use of local pitch peaks on words of particular importance or newsworthiness in an utterance;

The universalistic explanation captured relevant common features pertaining to the organization and functioning of pitch in language, but at the same time showed to have limits in the explanation of the relation between intonation and the communicative function in language and, in addition, showed to have significant limits in the description of the organization of intonation across different languages.

Firstly, within each language, there is no 1-1 correspondence between contours and pragmatic functions. One contour can be associated with different contexts of use. Moreover, different "shades" of the same pragmatic interpretation can be expressed by qualitatively different contours. Secondly, the universal approach fails to explain why in some languages a final declining contour it is reported as a prosodic marking of polar questions (D'Imperio & House, 1997; Savino, 2012), violating the assumption that a declination always corresponds to the notion of finality and completeness. The rather intuitive association of the final rising movement with the turn-taking function in conversation is also probably not universal, or at least requires a further detailed description of phonetic implementation (often it appears that the final rise is somewhat shifted in an earlier position within the prosodic unit, e.g., D'Imperio & House, 1997).

Contrary to the holistic approaches to the intonation description, the AM approach assumes that the abstract tones also have the status of "morphemes." In this way, the meaning can be composed as the sum of smaller discrete units and, in the same way, the tune constitutes the combination of discrete phonological representations (von Heusinger, 1999).
The meaning of a tune can thus be conceived as a combination of PAs, PhAs, and BTs. The meaning of each single unit is related to the node of the prosodic hierarchy to which they are attached and therefore granted by the hierarchical organization of the phonological domain (Pierrehumbert & Hirschberg, 1990). The phonological domain and its association with text determine, therefore, the pragmatic function of the contour.

The open question remained, therefore, to disentangle which aspects are shared across languages, and what aspects are linguo-specific. This has been addressed by Gussenhoven (2002). In Gussenhoven perspective both the biomechanical aspects and the phonological representations belonging to the grammar of each different language converge in a complex picture. The investigation of intonational meaning has to consider both perspectives. The biomechanical approach is therefore reorganized in 3 mechanisms, that the author defines as "codes" and that are assumed to be related to the act of speaking, therefore valid across languages. The previously described constraints generated by the mechanical properties of the vocal tract and by the act of phonation, as well as the segmenting of speech in function of breath groups, belong to what Gussenhoven defines as the Production Code.
The Frequency Code is the description of the relationship existing between the dimension of the larynx and of the vocal tract, and the rate of vibration of the vocal chords. The larger is the dimension of the larynx and of the vocal chords, the slower they vibrate, resulting in an overall lower range of pitch modulation. The perception of overall higher voices' pitch ranges is, thus, associated to the presence of a speaker (or animal in animals' vocalizations), potentially, physically more powerful and aggressive than a speaker displaying overall a significantly higher pitch range. In human beings, this mechanism indicates for example whether the speaker is an infant, a young adolescent or an older person, and helps disambiguate female and male speakers (Ohala, 1994).
The Effort Code indicates the effort that a speaker puts on the production of speech. Putting more effort in the production results in more precise articulatory movements and a higher pitch variation. The Effort Code has been generally put in relationship with the intention to convey emphasis to speech, associated with the span of modulation. A higher span of movements can be associated to a specific intention of the speaker to signal the relevance of the content uttered with an extremely high variation or aligned to a peak (Wichmann,

House, & Rietveld, 1997). The Effort Code plays a relevant role in the developing of models integrating the prosodic modulation with the information structure of discourse, that will be addressed in the section 1.2.2 and Chapter 4.

In Gussenhoven perspective, the 3 codes then have to deal with the grammatical organization of the phonological system in each language, and therefore they can be subverted if the specific grammar predicts it (e.g., conversational turn exchange through low Boundary Tones). According to Gussenhoven, thus, the property of discreteness is granted by the organization of pitch modulation in each language. Language grants to pitch the organization in discrete entities and systematization of meaning. Gussenhoven 2002, p. 2:

> "Intonation is used to route the semantic contents of particular morpho-syntactic constituents to semantic categories of information status."

Adopting such a perspective, it is possible to investigate both:

- the structure of the discrete linguistic nature of an intonational category, how it is possible to form it, to identify it and to place it within a phonological system - keeping in mind that the phonetic implementation is most probably what carries a universal function bound to the anatomical structure and functioning of the vocal tract;

- the mapping of identifiable discrete units with the pragmatic functions in different linguistic systems – observing whether, within the discrete units, specific phonetic features that are universally explained with biomechanical reasons systematically occur across languages – the role of falling and rising contours.

I will address in the following paragraphs first the substantial contribution of the adoption of the AM theory to the description of the association between pitch contours and pragmatic functions in discourse, and in the following paragraph 1.3. I will add to the debate the evidence coming from research in perception that opens relevant questions on the cognitive nature of an intonational category.

1.2.2. The pragmatics of Discourse

Pierrehumbert and Hirschberg in 1990, introduced two important aspects connotating the context of use of PA: the role of what Pierrehumbert calls the "mutual belief" between the interlocutors; and the instruction to explicitly or implicitly link a specific prosodic unit to an entity present in the mutual belief, by means of the resolution of an inferential process, considering all the levels of information (phonological, morphemic, and semantic).
In their work of 1990, the two authors proposed a functional interpretation of rising and falling accents: they considered the shape of the tone that anchors to the metrically prominent syllable a criterion to map whether an expression was either present, absent, or inferable in the mutual beliefs of speaker and hearer. In particular, they associated the PAs composed by a rising movement (L+H) with perceptual saliency, and PAs composed by a falling movement (H+L) with the presence of inferable information. In addition, they associated the presence of an H* tone with the instantiation of mutual belief (Pierrehumbert & Hirschberg, 1990).

The proposal along the years merged with theoretical descriptions of discourse structure that developed in theoretical semantics (e.g., Chafe, 1974), pragmatics (e.g., Prince, 1981), and along with the comprehensive modeling of the structure of language (Chomsky, 1971). The analysis of discourse structure and its integration with intonational description converged in the years in two main orthogonal approaches: the referential and the relational approach (Avesani & Vayra, 2004; Gundel & Fretheim, 2004). The referential approach (Chafe, 1976, 1994) is the analysis of discourse that considers the accessibility of information conveyed by single referents, as well as their anaphoric and cataphoric relationships to other referents in discourse. The relational approach (Büring, 2012; Vallduví, 1991), on the other hand, is primarily interested in the intentions of the speaker to signal whether entire prosodic constituents have to be considered either: shared and present in mutual beliefs; or newly introduced in discourse for the first time; or newsworthy and relevant, because contrasting or correcting information already present in the mutual belief.
Despite the two functional interpretations of pitch proposed by Pierrehumbert & Hirschberg (1990) - the one linking referents across prosodic units in discourse, and the one conveying perceptual salience to a specific referent within a prosodic unit – and the related approaches to the analysis of discourse refer to two independent levels of analysis, they are tightly interconnected between themselves and with intonational movements, and unfortunately often overlapped.
In Chapter 4, I will present evidence from neuroscientific research that supports the view that the two ways in which intonation maps with discourse can be reconducted to two distinct cognitive processes. Thus, I intend to address them separately.

1.2.3. The referential approach

The referential perspective focuses on the link between the mental activation of a referent and its corresponding non-linguistic entity (Baumann, 2005; Chafe, 1974, 1976, 1994; Gundel & Fretheim, 2004). Under this perspective, one important mechanism supporting the discourse construction is the Given-New hierarchy (Avesani & Vayra, 2004; Baumann, 2005, 2006; Baumann & Grice, 2006; Chafe, 1974, 1976, 1994; Clark & Haviland, 1977; Gundel & Fretheim, 2004; Prince, 1981; Vallduví, 1991). The term "given" expresses the semantic and pragmatic accessibility of a referent and is opposed to the term "new" that signals that the referent is introduced for the first time in the discourse. At the cognitive level, the term "given," signals, therefore, a fully activated representation of the entity, and the term "new," that the referent is not yet activated when processed by the interlocutor.

Within the Given-New hierarchy, the initial idea dealing with only two mental states - "given" (completely activated) vs. "new" (non-activated) - has been reformulated as a continuum, that can be segmented (Chafe, 1994) or hierarchically organized (Baumann, 2006; Gundel, Hedberg, & Zacharski, 1993; Prince, 1981) in more than two levels.
Chafe (1994) proposed that at least 3 levels of mental activation should be considered: inactive, semi-active, and active that respectively correspond to 3 statuses of the specific linguistic referent: new, accessible, and given. When a referent is encountered during on-line sentence processing, its status is defined as function of possible activation in the mind of the hearer, and it is considered: given when its mental representation is already active; accessible when it shifts from a semi-active state to full activation; and new when it reaches full activation from a previous completely inactive state (Baumann, 2005, 2006; Baumann & Grice, 2006; Chafe, 1994). This theoretical framework is especially attractive for cognitive neuroscience of language since it allows to link the abstract linguistic discourse-level representations (accessibility) to psychological

constructs of *mental activation* that can be linked to specific cognitive processes during on-line sentence comprehension.

Given and active referents are *contextually salient*, meaning they are easily retrievable from memory by both interlocutors and likely to continuously maintain their activation along the discourse development. At a functional and linguistic level of the discourse, the information is (considered to be) easily retrievable by both interlocutors as a function of different possible linguistic or non-linguistic facts: the referent has been previously mentioned directly or indirectly; its content is considered common world knowledge between interlocutors; a non-linguistic referent that is clearly associated with the linguistic one is physically processable in the environment at the time of the utterance. Given information is thus considered *identifiable* (Chafe, 1994; Baumann, 2005), meaning that it is possible during on-line processing for the hearer to anchor current information to a unique previously mentioned or implied antecedent, generating an anaphoric relation between two referents (Clark & Haviland, 1977). Given information is therefore assumed to be central in discourse, easily retrievable, and that can always be anchored back to an antecedent.

Adopting the Given-New hierarchy, Givenness is expressed at different degrees.

Textually given linguistic referents can directly anchor to the same identifiable and contextually salient antecedent in the immediately preceding portion of discourse; situationally Given referents can be directly identifiable within the extra-linguistic processable environment (Baumann & Grice, 2006).

Nonetheless, a referent can be identifiable even if not directly previously mentioned or explicitly present in the extralinguistic context. An anaphoric link can be established on the basis of semantic/pragmatic inferences (Baumann, 2005; Baumann & Grice, 2006; Lambrecht, 1996; Prince, 1981). A referent is connected to a set of related mental representations, and each of these connected entities is likely to receive a certain degree of activation during the processing of the utterance. Within the referential framework, these connected entities are defined as semi-activated, and they represent (more) accessible information (than completely unrelated new ones).

Typical examples of discourse structures instantiating this phenomenon are the Scenario and Schema settings (Sanford & Garrod, 1981; Tannen, 1979) where a single word corresponding to a complex set of objects (e.g., restaurant, office, station, classroom) possibly spreads activation to referents known to be part of the set (e.g., waiter, secretary, train, teacher). Vice-versa, an element of the set (e.g., a teacher) can spread activation to the scenario that typically contains it (e.g., a classroom).

In these cases, referents co-activated by the scenario are made accessible by bridging inference (Clark, 1977). A paradigmatic example of semi-active referents that will be at the center of the perception experiment in Chapter 4 are whole-part relationships (Girju, Badulescu, & Moldovan, 2006).

Whole-part (or vice versa part-whole) relationships are characterized by two referents: Y & X, where X is part of Y (Girju et al., 2006) and by the following axioms (Simons, 1987, 1991):

1) Existence – if X is part of Y, then X & Y both exist;
2) Asymmetry – if X is part of Y, then Y is not part of X;
3) Supplementarity – if X_1 is part of Y, then Y has a part X_2 disjoint from X_1:
4) Transitivity – if X is part of Y, and Y is part of Z, then X is part of Z;
5) Extensionality – if both Y & Z are composed by the same X_i then Y & Z are identical;
6) Existence of mereological sum – for any N of X_i it exists a Y that consists of $X_{i;}$

Winston, Chaffin, & Herrmann (1987) define 3 main ways to describe whole-part relationships:

Firstly, the relationship of a part with its whole can be established by functionality, as in "cup-handle" where the referent "handle" serves a specific function, and within the object, is placed in a specific position. Secondly, whole-part relationships are homeomerous when the parts are identical and of the same kind of the whole (e.g., pie-slice) and non-homeomerous when not (wheel-bike); and finally, parts can be separated from their whole (cup-handle) or not (alcohol-cocktail).
Adopting these 3 criteria, Winston et al. (1987) classify whole-part relationships in 6 types:

		Relation Properties		
Whole-Part relations	Example	Functional	Homeomerous	Separable
Integral Object - Component	Cup-Handle	+	-	+
Collection - Member	Deck - Card	-	-	+
Mass - Portion	Salt - Grain	-	+	+
Object - Stuff	Cocktail - Alcohol	-	-	-
Activity - Feature	Shopping - Paying	+	-	-
Area - Place	Desert - Oasis	-	+	-

Table 1.2.1 Whole-Part relationships - adapted from Winston, Chaffin, Hermann (1987) p.421

1.2.4. The prosodic marking of the status of a referent

In the previous section, the biomechanical universalistic description of the meaning of intonational contours has been described introducing the concepts of Frequency, Effort and Production codes (Gussenhoven, 2002). In particular, the Effort and Production Codes highlight the function of exceptionally high peaks, intentionally aligned in specific positions within a prosodic unit, that contrast with the naturally declining contour linked to the act of phonation and, thus, that can be linguistically interpreted as a voluntary signal of relevant information in the content.
By integrating the formalization of the Given-New continuum, expressing the degree of accessibility and activation of a referent, with the meaning that can be associated to the presence of declining or exceptionally rising contours, it is possible to formalize a description that associates the informational status of a referent with the choice of a specific pitch modulation signaling its degree of activation. Such a description allows decoding through the intonational contour whether a referent corresponds to identifiable contextual salient information or on the contrary to completely newly added information.
The regular presence of a link between specific pitch contours with the degree of activation of a referent has been extensively investigated in Germanic languages (English, Dutch, German). The main investigated principle is the one linking the newsworthiness and informational relevance of a referent with the presence of an acoustically salient pitch contour. Reversely, information that is already contextually salient and active in discourse is assumed to be identifiable and processable without the need to be prosodically marked with an exceptional peak; thus, it can be uttered with low levels of pitch. The generalization of this mechanism results in the association of given information with pitch patterns lacking perceptual prominence that are said to be deaccented (Cruttenden, 2006). The investigation of the prosodic marking of different degree of Givenness in Germanic languages highlighted that Given information is, indeed, systematically deaccented so that the phenomenon can be considered a typological property of these languages (Swerts et al., 2002, Cruttenden, 2006).
Nonetheless, accessible information, such as the previously described Scenarios settings and Whole-Part relations, does not convey either completely new information, nor completely given. On the contrary, an

inferable referent takes advantage of the contextual saliency that is established by an (indirect) anchor with previously processed information and, at the same time, it enters into the discourse structure as a new independent referent (Burkhardt, 2006, 2007).
In English and German, accessible information is proposed to be marked by composite PA aligned to semi-active and inferable referents. The phonological structure of these composite PA appears to display properties that are analogically linked to the status of semi-activation at the level of discourse; it is in fact neither only a high-level tone (signaling new identities), nor only a low contour (signaling givenness), but on the contrary a mixture of the two.

1.2.5. Prosodic marking of referential Givenness in German

In German, a long tradition of studies has investigated the prosodic marking of referential Givenness (Kohler, 1991; Grice & Baumann, 2001; Baumann & Hadelich, 2003; Grice et al., 2005; Baumann, 2005, 2006; Baumann & Grice, 2006; Rohr & Baumann, 2010; Rohr & Baumann, 2011).
Kohler (1991) proposed the early peak contour, which is a movement with a high peak preceding a lower valley that is aligned to the accented syllable. Within Kohler's framework, the early peak contour is opposed to the medial peak that aligns at the center of the prominent syllable, and it is the preferred PA for new information. Baumann & Grice (2006) tested experimentally the perception of three distinct PAs: H* (high level tone, or peak aligned to the metrical head), L (low pitch contour corresponding to Deaccentuation) and H+L* (Falling pitch contour) on different types of accessibility finding the preference patterns reported in Table 1.2.2

Type of Accessibility	Pitch Accent Type Preferences	Preference for Deaccentuation of Target Referent (2nd one)
Converseness	Deaccentuation > H+L* > H*	Higher Preference ↓ Lower Preference
Part-whole	Deaccentuation > H+L* > H*	
Synonymy	Deaccentuation > H+L* > H*	
Hyponym-hypernym	Deaccentuation > H+L* > H*	
Hypernym – hyponym	Deaccentuation > H+L* > H*	
Textually displaced	H+L* = Deaccentuation > H*	
Whole – part	H+L* > H* = Deaccentuation	
Scenario	H+L* > H* = Deaccentuation	

Table 1.2.2 Results from Baumann & Grice (2006)

Rohr & Baumann (2010), in a following production experiment, investigated whether accessible information was uttered with a preferential PA or deaccented. Both experiments find a hierarchy of preferred PAs that can be interpreted according to the concept of Effort Code by Gussenhoven (2002). In particular, results highlight that specific types of accessible information are marked differently. Within the different types of accessible information tested by Baumann & Grice (2006), Whole-part relationships show a clear preference to be marked with H+L* and, in order, both high medial peaks (H*) and Deaccentuation are unpreferred. In particular, it can be noted that whole-part and part-whole relations display reversed preferences of associations of PAs. The Authors explain these findings highlighting that in part-whole relationships the subordinate expression (part) precedes the superordinate expression (whole) establishing in first instance the superordinate concept, that therefore following the Effort Code principle, does not require to be marked with a PA when encountered afterward. Reversely, the whole-part processing sequence instantiates

a sub-specification of the superordinate concept, guiding the attention of the hearer towards a specific subordinate concept (part). Under the Effort – Code principle this justify the modulation of pitch in association with the newly introduced referent. The apparently incongruent result for hypernym-hyponym relationships, for which a pattern similar to whole-part relationships should be expected, is explained by the Authors as a function of the stimuli uttered in the experiments where the referents expressing the hyponym concepts were always definite NPs. This might have induced an interpretation of the hyponym concept as a co-referential synonym of the hypernym concept, leading to a similar preferential pattern of intonation.
It appears, therefore, that in German Intonation signals, and is strongly bound, to the degree of activation of a specific linguistic referent. It also emerges that native speakers of German systematically rely, consciously or unconsciously, on this relationship between the two linguistic domains, when constructing a representation of discourse, and when establishing explicit and implicit anaphoric links between two linguistic referents.

The prosodic marking of referential Givenness will be at the center of the study presented in Chapter 4. In the introduction of the specific chapter, I will furtherly address the available evidence at the neurophysiological level of the processing of the perceptual integration of the intonational contour with the degree of activation of a referent. In particular, the study will be centered on the perception of congruent and incongruent pitch contours aligned to semi-active referents, within a whole-part relationship in discourse.

1.2.6. The relational approach – the notion of Focus

The relational perspective originated from the work of Halliday (1967), who considered from the beginning both the information structure of discourse and the intonational level. In his framework, the discourse is decomposed in two main parts – information units – that are in relation with intonation by means of "tone groups." An information unit is assumed to be uttered within the boundaries of one tone group and can be of two types: informative or non-informative. In this way, the discourse is partitioned in a part conveying "what the utterance is about" (non-informative) and a second one that adds information about the first one (informative). In Halliday's proposal, a substantial role in the partition is given by word order, in fact, the non-marked structure of discourse information predicts that the non-informative part precedes the informative one and that the presence of a PA is associated with the most informative part.
On the basis of the concept of informativeness, several theoretical approaches have developed generating a complex set of concepts, often overlapping. The source of confusion is the presence of different assumptions and implications due to each proposal and to the use of interchangeable terms, associated to different concepts within each proposal, that overlap when an analysis is done considering more than one theoretical approach.
A general parameter of informativeness would also include the opposition between given and new information that is at the base of the referential approach. The reasons for separating the relation between given-new information is that it specifically pertains the degree of cognitive activation of the single referent, while the following approaches are centered on the intention of the speaker to highlight a specific portion of discourse, independently from its degree of activation. This is done introducing the concept of focus, that is the element that allows to a partition of discourse to stand out from the rest of the information.
The relational approach considers the intentions of the speakers and its belief about the current knowledge of the hearer on an independent orthogonal level respect to the Given-New opposition. The speaker can

intentionally choose to highlight the discourse elements that he/she believes are the most newsworthy for the hearer, independently from their degree of activation.

Dichotomies	
Theme-rheme	(Ammann, 1928; Daneš, 1994; Firbas, 1964; M. A. Halliday, 1967; Mathesius, 1936);
Topic-comment	(Reinhart, 1982);
Topic-focus	(Hajicová, 1980; Hajičová, Sgall, & Skoumalová, 1995; Partee, 1998; Sgall, Hajicová, & Panevova, 1986)
Presupposition-focus	(N. Chomsky, 1971; Jackendoff, 1972);
Background-focus	(M. A. Halliday, 1967; Jacobs, 1984; Kuno, 1978);
Link-Focus-Tail	(Vallduví, 1991);

Table 1.2.3 Relational structures

From the Table 1.2.3 above, it is possible to observe how the early conceptualizations strongly relied on the primary partition, based on the intuitive idea, that what is not informative appears first in the sentence and serves as an anchor point, and afterward, the following partition completes the structure, attaching to it. This type of representation directly associates the idea of givenness with the concept of theme and the idea of newness with the following rheme partition of discourse.

In Halliday's terms (1967b, p.212), this is also expressed by the position of the discourse partitions. The theme is: "what is being talked about, the point of departure for the clause a message" and "what comes first in the clause." Therefore, following Halliday's approach, in the 3 following examples what is considered theme varies in each sentence:

1. [John]$_T$ [saw the play yesterday]$_R$;
2. [Yesterday]$_T$ [John saw the play]$_R$;
3. [The play]$_T$ [John saw yesterday]$_R$;

The topic-comment dichotomy shares conceptual properties with the theme-rheme relation, but also it introduces the concept of aboutness and allows the less informative part of the discourse not to be necessarily displaced at the beginning of the utterance.

The concept of focus introduces a deeper level of complexity. The focus element was already present in Halliday's formalization as the informative subpart of the rheme partition. The concept of focus introduced by Chomsky (1971) and Jackendoff (1972) in a generative grammar perspective is a hypothetical construct based on the sematic-pragmatic notion of new information, having a syntactic correlation expressed through a syntactic focus-feature F and prosodically marked by a PA at the prosodic level. The key point is that the notion of informativeness within the relational approach is not a property of the referent and its degree of activation, but it depends on the contextual information. A portion of discourse is focused with respect to a complementary portion that lies outside its scope. The approach is "relational" because discourse related.

Thus, an utterance can be partitioned in focused and "non-focused" elements depending on the preceding context, determining the scope of focus. The scope of focus can shift from including the entire utterance, to specifically highlight only a single word (Broad vs. Narrow). The type of focus signals the relation of the semantic content of the element in focus with the preceding context, specifying whether the information has to be simply added to the previous representation of discourse (informative), or on the contrary signals

the presence of two elements in contrast (contrastive), or ultimately corrects the already available representation of discourse (corrective).
The following table shows an overview of the main focus types, their link with the preceding context, and their scope within the utterance:

Context	Utterance	Focus type
What happened?	[Mary gave a present to John]F	Broad Focus
Who gave a present to John?	[Mary]F gave a present to John	Narrow Informative Focus
What did Mary give to John?	Mary gave a [present]F to John	Narrow Informative Focus
To whom did Mary give a present?	Mary gave a present to [John]F	Narrow Informative Focus
What did Mary do?	Mary [gave a present to John]F	Narrow Informative Focus
What did Mary do to John?	Mary [gave a present]F to John	Narrow Informative Focus
Did Mary give a present to Peter or John?	Mary gave a present to [John]F	Contrastive Focus
Mary gave a present to Peter.	Mary gave a present to [John]F	Contrastive Corrective Focus

Table 1.2.4 QUDs and Focus Types

The context mechanism guiding the position and the scope of the focus feature in the target sentence is the Question Under Discussion (QUD)(Büring, 2012). Question and answer have to respect a congruence pact, and therefore the element in focus always needs to convey the relevant information requested by the context. Once the semantic content of the utterance is determined to reply to the question successfully, the syntactical and phonological structure of the utterance is organized by the hierarchical metrical structure, and by the principle that, an element in focus requires to be prosodically marked. This procedure, thus, determines that the pragmatic context sets a criterion of relevance and informativeness and once it is set, the PA positioning, as well as the discourse structure, are predictable on the basis of the focus projection rules (Chomsky & Halle, 1968).
The presented theoretical formalization is the one adopted by the "Focus-to-Accent" approach (Gussenhoven, 1983; Ladd, 1980; Janet Breckenridge Pierrehumbert, 1980; Selkirk, 1984) within AM, in order to map the relation of PA in each linguistic system. It allows to disentangle in which context a specific PA appears in which language, and it serves the function captured by the semantic dimension proposed by Ladd (1996).
Moreover, it is now possible to observe why the given-new formalization (referential approach) is orthogonal to the dichotomies adopting the concept of focus (relational approach). In the following example the QUD sets as information shared by the interlocutors the referent Jones. In referential terms, Jones is then textually given information. In relational terms, on the other hand, the pragmatic context commands that Jones corresponds to the focused element in the answer since it represents the relevant information. In the reply, thus, the referent Jones is focused, and therefore accented, despite the fact it is given information.

4. A: Who did Jones' father recommend for the job?
   B: He recommended JONES for the job.

Dahl (1974) summarizes the 4 abstract entities organized in two orthogonal levels in the example below:

5. A: What does John drink?

B: $[\text{John}]_T$ $[\text{drinks beer}]_C$ ⟶ Topic – Comment;
B: $[\text{John drinks}]_B$ $[\text{beer}]_F$; ⟶ Focus – Background

The last model relevant for this work is the one proposed by Vallduvì (1991).
Vallduvì (1991) introduces a further specification of the notion of discourse structure. In particular, he defines the information packaging representing the whole conceptual framework as a set of instructions that the speaker conveys to the interlocutor (p.11):

> "Information packaging is taken to consist of a small set of INSTRUCTIONS with which a speaker directs a hearer to retrieve the information encoded in a sentence and enter her/his knowledge-store. The purpose of information packaging is precisely to optimize the entry of data into the hearer's knowledge-store."

In Vallduvì 's formalization, the concept of information is set apart from the semantic level and from the relation between speaker and hearer, and expressed through a symbolic formula:

$$I_s = p_s - K_h$$

where I_s stands for "Information of the sentence," p_s corresponds to the "propositional content," and e K_h represents *the knowledge of the hearer*.
The formula states that in a sentence, the information is the delta emerging, subtracting the knowledge of the hearer from the overall semantic content present in the utterance. What remains is the actual information.
In these terms, this formalization merges the concept of focus representing the most informative part of the utterance in Halliday's terms, with the partition of the discourse that in the generative approach is the complementary element of the focus that is the non-informative part of the utterance in the background.
The propositional content is partitioned in a Focus-Ground dichotomy. Also, in Vallduvì's system, the Ground is furtherly subdivided into Link and Tail components. The Ground, intended as a complementary part of the Focus, is considered "equivalent to the concept of presupposition, open-proposition, Prague-topic, and background" (Vallduvì, 1991, p. 58). It functions, within this view, as a set of instructions pointing to the hearer, where information must be added (informative focus sentences) or substituted (contrastive focus sentences).
The link, as a subcomponent of the Ground, has an analogous function to the sentence-initial topic, in topic-comment sentences: "the link directs the hearer to a given address in the hearer's knowledge-store" (Vallduvì, 1991, p.59) and must be sentence-initial. Finally, the Tail is considered complementary to the link, and signals right-detached partitions of the ground that add further specifications to the link component.

1.2.7. Plasticity

A source of confusion in the study of typological differences in the prosodic marking of Discourse Structure comes from the comparison of evidence elicited in different languages either through the Referential or Relational approach, or a mixture of the two. The literature on the intonational system of Germanic languages, especially on Dutch and German, provided a detailed description of the prosodic marking of referential statuses (Given-New), as described in paragraph 1.2.3. The literature on Romance languages, on the other hand, has focused on the description of the prosodic marking of several focus types in different syntactical constructions (Focus-Background), that will be addressed in the following section 1.2.8.

A particularly active debate is the one concerning an assumed typological difference between the prosodic marking of information in Germanic and Romance languages. In his dissertation and a following study, Vallduvì (1990, 1991) introduced in the literature the plasticity parameter within his own specific model of discourse structure – the information packaging (classifiable as a relational approach) – and described the mapping of intonation and information structure in Catalan, highlighting a systematic difference with the mapping occurring in English. In Vallduvì's description, plasticity is a factor that can assume two levels (Vallduvì, 1991; p. 295)

> "[+plastic]: intonation contour may be molded to attain the togetherness of focus and prominence (English).
> [-plastic]: intonation contour may not be molded to attain the togetherness of focus and prominence, which must be attained by other means (Catalan)"

Plasticity came to highlight a specific property that seems to occur across Germanic languages, namely the possibility to maintain unvaried the syntactical structure of an utterance – that guides the position of the metrical head and of the Nuclear Pitch Accent (NPA) – while shifting the focus position away from the default metrical head, in function of the type information requested by the preceding context – the Question Under Discussion (QUD) (Büring, 2011). The presence of a focus element early in a sentence, with respects to the metrical head positioned on the right end of the utterance, generally causes the presence of an intonational peak aligned to the focus element, and as a consequence the Deaccentuation of the portion of the utterance still in head-position. On the contrary, the property of being non-plastic came to indicate the absence of this mechanism, and the stronger constraint imposed by the position of the metrical head. In languages like Catalan and for extension Romance languages, it is observed in a strategy that overlaps the position of the focus element with the metrical head position, and the consequent rearrangement of the sentence structure in cases the context forces a different organization. The consequence is that, for what pertains Romance languages, the occurring of Deaccentuation is highly debated and sometimes studied mixing the two approaches, so that it is not clear yet what is the guiding principle behind it.

When performing a comparative analysis of the systems, it becomes highly probable to associate the prosodic marking of referential New information, to the prosodic marking of the Focus element within the relational approach. This is due to the fact New information is naturally placed in focus-position since it represents the best instance of newsworthy and informative content in discourse. It follows, that non-plasticity by extension merges the resistance of Romance languages to deaccent information in nuclear position with the prosodic marking of the information status of a referent independently from its degree of activation (Given-New continuum). The presence of Deaccentuation due to a shift of a Narrow Focus earlier in the sentence can overlap with the Deaccentuation phenomenon described within the referential approach, where completely active, given, identifiable, and contextually salient information can be fully accessed and processed without the presence of a specific prosodic marking highlighting it. Nonetheless, the reasons and the mechanisms at the base of the two Deaccentuation processes, are different and relating to two orthogonal and independent perspectives.

Plasticity, also, is not only studied focusing on Deaccentuation and on the position of focus elements, but also by observing whether the syntactic organization of a sentence is linked to the expression of different focus types. This is the perspective adopted by Face & D'Imperio (2005), who proposed to overcome the dichotomy and the adoption of a rather continuous modulation of the factor "plasticity," showing that within Romance languages the expression of focus is declined in different ways. The proposal is grounded on data on Castellan Spanish and Neapolitan Italian signaling differences in the focus realization in the two languages. In particular, they highlighted that Italian could grant a prosodic distinction to Broad and Narrow Focus realizations, respectively uttered through H+L* and L+H*, while in Spanish this distinction is absent and both Broad and Narrow Focus sentences are uttered with L+H*. Thus, Italian, according to the authors, shows the property to be more plastic than Spanish since it can signal a discourse structure distinction through intonation alone, with a reorganization of the content in the sentence. They proposed, therefore, a formalization in which languages lay on a continuum, in which on one end lie languages favoring "intonation-only" strategies, corresponding to the original definition of plastic languages, at the opposite, there are languages strongly demanding for a syntactic re-organization in order to express different focus-background structures, corresponding to the original concept of non-plastic languages. Within this framework, Italian has been positioned closer to the latter end but displaying significant differences from Spanish.

Nonetheless, the position of D'Imperio on the Narrow Focus marking in Neapolitan, on the basis of the same data, varied along the years in 3 different studies (D'Imperio, 1997; D'Imperio & House, 1997; Face & D'Imperio, 2005). Broad Focus sentences are consistently reported to be marked through H+L*, while Narrow Focus is marked as starting assumption with H* and revised in H*+L on the basis of the evidence of the study (reported in paragraph 1.3.6.) on the role of alignment in the perception of statements and answers. The main evidence reported in D'Imperio & House (1997) is the fact that in Neapolitan Italian statements appear to be categorically identified not by the presence of a peak alone, also present in the marking of question, but by the falling movement through the vowel, opposed to a rise signaling the presence of a question. The authors in this paper concluded proposing: "a unified analysis of statement pitch accents as consisting of an HL sequence independent of breadth of focus, where H+L* will be used as a label for broad focus utterances and H*+L, instead of monotonal H*, would then be employed for labeling narrow focus ones." (D'Imperio & House, 1997; last page). An analogous description is also reported in D'Imperio (1997; p.23) describing Narrow Focus elements as marked with: "a nuclear pitch accent that is acoustically more salient and are characterized by a H*+L nuclear accent." In addition, in D'Imperio (1997) the recognition of focus type through a perceptual procedure requiring to match a perceived target statement, in which the focus structure shifts from Broad Focus to Narrow Focus highlighting either the subject, the verb, or the object, is found significantly more accurate compared to recognition of the same structures embedded in questions, but displaying high variability in the correct identification of the scope of Broad Focus structures. The same Broad Focus and Narrow Focus structures are assumed to be identifiable in Neapolitan Italian, marked by falling (H+L*) and a rising accent (L+H*) in Face & D'Imperio (2005). Despite that a reanalysis of a label of pitch contour is frequent in the study of intonation, the label for Narrow Focus in Face & D'Imperio (2005) marks a rising contour, therefore it might appear in contradiction with the previous considerations stating that statements independently from the type of focus are always marked with a falling HL sequence.

Since the proposal to revise the entire concept of plasticity is grounded on the assumed difference between two varieties of Spanish and Italian, and the interpretation of the data concerning Italian has been repeatedly revised, it is not entirely clear how non-plasticity is expressed within different Romance languages, and how the expression of different focus types interacts within one comprehensive model with the possibility in Romance languages to deaccent given information and the interaction of the degree of activation of information (Given-New continuum) with the positions of prominent features within the metrical structures. I suggest that more data are necessary, disentangling the role of each factor in a comprehensive description in order to explain at a deeper level the functioning of plasticity.

1.2.8. Prosodic marking of Discourse Structure in Italian: Focus types and Deaccentuation

The description of the phonological system of Italian can be harder to provide compared to the one of other languages (e.g. English, Dutch or German) given the presence of different regional varieties, which can diverge among each-others in some main phonological and phonetic features, and also in the way they interact with the structuring of information in Discourse. A comprehensive attempt to furnish a common phonological inventory is provided by Gili Fivela et al. (2015), described at the beginning of the introduction, paragraph 1.1.3, in which it also reported and summarized the available evidence in each variety of the link between intonation and Discourse Structure. I report in the following paragraphs the main evidence that will be useful to understand and interpret the experimental manipulation proposed in Chapter 4, which deals both with the referential description relevant for the processing of intonation in native speakers of German, and the available data on the processing of discourse structure in native speakers of Italian, which is strongly centered on the relational approach.

In the literature addressing the expression of discourse structure in Italian, there is consensus that broad focus and narrow focus, as well as Informative and Contrastive sentences, can be differently marked across varieties (Grice, 1995; Avesani & Vayra, 2000; D'Imperio, 2001, Grice et al., 2004, Avesani & Vayra, 2004; Face D'Imperio, 2005; for a recent review see Gili Fivela et al., 2015). Despite the plurality of results, there is a rather good agreement on the fact that broad focus sentences are frequently associated with H+L* and that, in most of the varieties, L+H* is systematically associated with contrastive focus information (Gili Fivela et al., 2015).

Pertaining to the prosodic marking of information status, several early studies are in line with the original dichotomy highlighting a lack of use of Deaccentuation and the favoring of syntactical reorganization.
In a read speech study Avesani (1995) found an association of a low flat F0 contour, annotated as L*[2], with discourse external entities such as vocatives, and in a following broadcast speech study (Avesani, 1997) the same L* accent has been found to be associated with given referents. More recently, Avesani & Vayra (2005) recorded semi-spontaneous speech in a setting in which two L1 Italian (Roman variety) speakers (who could not see each other) had to describe each other slightly different images. Co-referential identities in the dialogue were deaccented with a proportion of 0.07 while the majority (0.93) of the co-referential productions displayed an H* PA. Not only the low proportion of Deaccentuation is in line with the traditional dichotomic hypothesis, but also the Authors note that the few cases of Deaccentuation are likely to be motivated by structural metrical-syntactic constraints.

Cruttenden (2006) presented a cross-linguistic study by comparing productions of 10 dialogues, translated and recorded in 12 languages by native speakers of each language. The overall pattern of Deaccenting and Re-accenting textually given information interestingly shows to be in line with the Face & D'Imperio (2005) continuity proposal. The observed languages appear to be ordered from an extremity highly deaccenting given expression, to an extremity highly re-accenting the target expression. With respect to Italian, textually Given elements were surprisingly associated with Deaccentuation with the large and consistent proportion

[2] By labeling a low flat F0 contour with L* rather than Deaccented, Avesani (1995, 1997) informs us that this contour, despite being low, is aligned to the metrical structure and thus has to be considered a marked PA. This does not allow to associate Deaccentuation and L* with a same phonological entry and a same pragmatic function. However, within a shallower interpretation of pitch contours (e.g., based on the analogical Effort Code framework) one has to consider that a low flat F0 contour may receive functional interpretations that are not entirely unrelated to Deaccentuation.

of 0.60. Interestingly, in the remaining 0.4 proportion of cases, in which given referents were associated with a PA, the re-accented element appeared in a structure that might show analogies to the ones originally adopted as default example by Ladd (1996, 1998). In Ladd's view, in Italian, a rule is present that blocks PA moving within a syntactic unit when the shift in the position of the PA is guided only by contextual information structure. Although the overall cross-linguistic patterns are in favor of the continuity hypothesis, this latter observation is also partially in line with the idea that the restricted use of Deaccentuation in marking given referents in *"non-plastic"* languages is likely to be motivated by metrical-syntactic rules.

Typological differences in marking Givenness have been studied cross-linguistically by analyzing the production (and perception) of intonational contours associated to the information status of a referent during card-games or following a computer-routine (Swerts et al., 2002; Rasier Hiligsmann, 2007; Avesani et al., 2015). In these experiments, the production of simple NPs was guided by a pre-set sequence that manipulated the change of information status of the referents along the task. The sequence was constructed by creating different relationships among successive NPs, so that at the beginning of the procedure the first NPs conveyed discourse's new information both through a noun and an adjective (New-New, NN), then alternating, the noun is varied but keeping the adjective conveying textually given information (Contrastive-Given, CG), or contrasting only the adjective (Given-Contrastive, GC), or contrasting both noun and adjective (Contrastive-Contrastive, CC).
Cross-linguistic comparisons contrasted Dutch with both French and Italian (Swerts et al., 2002; Rasier & Hiligsmann, 2007), finding that, in L1 Dutch data, contextually new referents were always accented and given information deaccented. In French and Italian, Rasier & Hiligsmann (2007) and Swerts et al. (2002) found in the production of Romance language speakers the presence of a bridge accent aligned to the whole NP independent from the information status of the referent.

Rasier & Hiligsmann (2007) have also investigated SLA patterns of acquisition of the prosodic marking of information statuses. They compared the acquisition of the prosodic marking in Dutch as L2 by L1 speakers of French, and in French as L2 by native speakers of Dutch. In the first study, the Authors found in the production of L2 speakers of Dutch the presence of the same bridge accent along the NP across all experimental conditions, and only a proportion of Deaccentuation of Given referents of 0.25, compared to the 0.75 proportion in L1 speakers. In the production of L2 speakers of French, L1 speakers of Dutch showed a similar behavior to L1 speakers of French.

Pertaining Italian, SLA patterns of acquisition of prosodic marking of referential statuses have been exploratively investigated by Avesani et al. (2015) with a similar card game paradigm as in Swerts et al. (2002). The participants were both L2 adult learners of German (plastic) and of Italian (non-plastic), sharing the opposite language as L1. They found a systematic link between Deaccentuation and Givenness in German productions by L1 German speakers, while Italian productions by L1 Italian speakers showed various patterns of intonational contours not systematically related to the Information Status of the referent. Data on L2 speech, by the same participants, showed that L1 German speakers produced the appropriate pattern of prosodic realization required by structural constraints of Italian (mainly H+L*) but they did not produce the variety of contours produced by native Italians in their L1. On the other hand, the L2 German production of L1 Italian speakers inappropriately reproduced the same plurality of PAs they used in their language in L2, without exhibiting the typical association of Deaccentuation with Given information that is typical of German.

All the 3 studies (Swerts et al., 2002; Rasier Hiligsmann, 2007 (L1 & L2); Avesani et al., 2015) that used this fixed sequence of NPs to elicit productions of contrastive and given information can be compared, and they homogeneously found that Dutch and German speakers systematically deaccent given information. On the contrary, French and Italians in the same experimental conditions do not to systematically deaccent given information and hardly acquire it in L2 (Rasier & Hiligsmann, 2007; Avesani et al., 2015).

By summarizing, it is not possible to affirm that Italian (as well as other Romance languages) never uses Deaccentuation in association with given information (e.g., in Cruttenden, 2006, reported a proportion of 0.6 cases of Deaccentuation of given referents). It also appears that the analogical Intonation strategy to reduce discourse relevance and phonological saliency of given information by deaccenting (Effort Code hypothesis) is modulated by linguo-specific properties. In particular, this strategy is at least much less systematic in Italian and other Romance languages than in Western Germanic languages, and differences among Romance languages have also been underlined (Cruttenden, 2006; Face & D'Imperio, 2005).

I think that the reasons why Italian shows a less systematic link between specific PAs and Information Status of a referent can be multi-faced in nature, by entailing: (a) stronger metric-syntactic restrictions in PA variations; (b) a preference in systematically associating specific PAs to relational (e.g., focus) rather than referential (e.g., Givenness) aspects of discourse structure; (c) a lack of stability of the association between Intonation and Information Status across different varieties of the language. This can also be compatible with proposals that suggest going beyond the dichotomic distinction between plastic and non-plastic languages (Face & D'Imperio, 2005).

1.2.9. Summary

In summary, the universal aspects of intonation that link with the organization of discourse information are linked to the physical configuration of the vocal tract and to the biological act of breathing. The biological constraints are responsible for the production of a non-marked pitch modulation that is configurated with the highest level of energy at the onset of the prosodic constituent and that naturally declines depending on the necessity to breath (Liberman et al., 1967). This mechanism is named the Production Code (Gussenhoven, 2002). Analogously, the dimension of the vocal tract as well as the rate of vibration of the vocal chords, universally convey information about the biological characteristics of the speaker, like age, sex, and probable size. These biological features associated to pitch are group under the term Frequency Code (Gussenhoven, 2002). A modulation reversing the non-marked declination pattern is a modulation that it is possibly linked to attentional mechanisms since the deviation becomes perceptually salient and can be captured by the attentional mechanisms of the perceiver. This mechanism is captured both by sound-symbolism perspectives (Ladd, 1996) and the Effort Code, in Gussenhoven description. The rising pitch contours given their perceptual saliency can be associated with turn-taking mechanisms (questions vs. statements), with the degrees of informativeness and referential accessibility (New vs. Given information), and with the communicative intentions of the speaker to grant perceptual saliency to part of the discourse, independently from its referential accessibility (Focus vs. Background).

On the other hand, the way the universal aspects of pitch are shaped into discrete, recognizable categories and aligned to text to convey specific pragmatic functions is language dependent (Gussenhoven, 2002). Thus, the association of a rising pitch, without considering its alignment, cannot be indisputably linked to the presence of a question, or the presence of new information across languages, and it cannot be assumed that the correspondence is always 1 to 1. The presence of a native language, thus, adds an arbitrary tune-text association reflected in different phonological inventories and different associations rules. This will also be addressed in section 1.5.

When considering typological differences between West-Germanic and Romance languages, a main difference reported by the literature is the debated notion of plasticity (Vallduvì, 1991; Face & D'Imperio 2005). The notion of plasticity concerns the way information is packed in discourse (Focus vs. Background modeling) and conveyed to the interlocutor mainly modulating pitch (plastic) or mainly reorganizing the content through syntactical movements (non-plastic). According to the literature, Romance languages appear to be non-plastic, and West-Germanic languages to be plastic. Nonetheless, non-plasticity has been proposed not to apply to Romance language homogeneously, but that on the contrary, the properties are modulated on a continuum, and Italian (a specific variety), for example, can be considered more plastic than Catalan (Face & D'Imperio, 2005).

On one side, the notion of plasticity calls into play the processing of elements in Focus vs. elements in Background, and the possibility to associate a perceptually salient peak to elements in Focus that is present in both groups of languages. Nonetheless, on the other side, it also calls into play the possibility to adopt a Deaccentuation strategy, reported in association with referentially given information within Germanic languages. Therefore, when focusing on perceptual and L2 acquisitional mechanisms involving these typologically different groups of languages, it becomes necessary to point out what the two mechanisms share, and where on the other hand, are distinguishable. In this section, therefore, I have highlighted the necessity to keep the two dichotomies apart, and the reasons why it has been challenging to do it.

Perceptually, both New information and elements in Focus are associated with the presence of a rise. The property of a rising contour to be perceptually salient belongs to the phonetic implementation linked to the biomechanical mechanisms; thus, it is worth to highlight that an element in Focus is always prominent within a prosodic constituent. It is possible to hypothesize that universally an element in focus links to the presence of attentional mechanisms, and this will be addressed in Chapter 4.

In this section, on the other hand, it is highlighted that the perceptual saliency associated with referential accessibility is granted by the semantic relation and information packaging across prosodic components. The communicative intentions of the speaker grant the perceptual saliency associated with elements in Focus. The distinction is relevant in perception, when focusing on the link between the sound processing and the pragmatic abilities of the perceivers, because the efficient processing of referential accessibility requires to infer the presence of a semantic relation between two elements. On the other hand, the processing of an element in focus requires to link an element acoustically prominent with a communicative goal, given the information packaging of the preceding content. The analysis of the literature suggests that these two aspects might have been overlapped in studies in the past, and in the aim of this work to consider them separately (Chapter 4).

1.3. The perception of speech: segmental and intonational categories

In the next section, I introduce the notion of category and its role within speech perception, in particular focusing on the perception of pitch. The debate around the notion of category is necessary since the literature on speech processing, and L2 acquisition at the segmental level, in most cases is grounded on the notions of phone and phoneme. In particular, I am going to present some of the most relevant theories, and effects classically found and established in the perception of segments (e.g., the Perceptual Magnet Effect and the Categorical Perception paradigm). Two main facts need be to debate in relation to the notion of category: i) what is its role within the perceptual system and how it emerges. It is worth mentioning that, there are also speech-processing models that sustain that the presence of discrete and abstract categories, such as the phoneme, are not necessary to efficiently process speech. One model arguing against the necessity of having discrete categories that I will introduce is the Polysystemic Speech Perception model (Polysp) (Hawkins, 2003). ii) whether intonation as complex phenomenon truly functions analogously to segments, therefore whether it is analogously perceptually organized in discrete categories, and whether they emerge in a similar way to segments.

It is necessary to address the topic about the perceptual nature of intonation, on one side, because by adopting the AM theory as theoretical structure, we adopt a system that assumes the presence of abstract discrete categories. On the other one, because given the fact that AM is adopted in L2 acquisitional studies, the mechanisms that are known to guide and regulate the acquisition of the L2 segmental categories have been along the years naturally adopted as a baseline to develop models about the L2 acquisition of intonation. Given the broader the association between the sound domain and the information packaging domain discussed in the previous section, it is useful to discuss what the actual presence of an intonational category implies within the sentence-level perceptual domain.

1.3.1. **The notion of Category**

The adoption of a system such as the AM theory, strongly distinguishing abstract representation from continuous phonetic implementation, as well as the cross-linguistic comparison of the way intonation is interpreted in meaningful representation, that presumably involves both levels of analysis requires to reason on what actually is an intonational category, and what aspects of an intonational categories are univocally identified across languages, opposed to the aspects that are linguo-specifically implemented. Where is the linguo-specific representation impacting? On which aspects of the intonational categories?
In an extensive overview focusing on the notion of category, Cohen & Lefebvre (2005) report at least 8 research areas that make use of the concept of category, having different implications from field to field. The concept can be used to refer to cognitive states and processes (Harnad, 2003; Harnad, Cohen, & Lefebvre, 2005), grammatical categories (Muysken, 2005), concept and empty concepts (Rey, 2005), their structural arrangements (Boster, 2017), the representation the brain makes of them (Hanson & Hanson, 2005), their implementation in artificial systems (Sowa, 2005). Categories are grounded in the organization of distinctive features in linguistics, at the same time are grounded in experience within the field of philosophy and embodied cognition (Barsalou, 2005; Lalumera, 2005) and so on, accessing several fields of science.
I address here the notion of categorization suitable in a cognitive and linguistic perception, that can explain specific properties related to auditory perception. There are two levels of analysis that sometimes are

combined in one line of reasoning creating confusions in the interpretation of the outcome. One problem is establishing the role of categories, what do they serve, and how they are formed, how they are put into relationship with one another and updated. A problem of a different nature is to investigate how categories are used once they are established to perform operations on the incoming signal. The first level of analysis focuses on the categorization process itself. The second one addresses the fact that once we have a set of available categories, we use them as sort of filter to process the real signal, and this is the level in which the research on Categorical Perception (CP) works. The research on auditory perception with the aim to extend the knowledge of linguistic processing began with the second level of analysis and the developing of the CP paradigm. The research on first language acquisition, being interested in explaining how language develops in the child is far more concentrated in understanding the first level of analysis, therefore in understanding how categorization begins in infancy. The research on second language acquisition of phonology has to deal with both levels, especially in late learners, since a phonological inventory, as well as other linguistic structures in the domain of morphology, syntax, semantics, and pragmatics, are formed, but the system is forced to actuate a process of restructuring of the available linguistic categories to introduce the ones functioning in the L2 system, not present in the native language, and to specify for existing categories working in both languages' different contexts of use. I explain therefore in the next sections what both levels imply.

Generally speaking, a category can be defined as the result of a categorization process. "Categories are kinds" (Harnad, 2005; p. 22). The members of a category share at least a common property that puts them into relation and they can diverge in minor properties, considered non-relevant for the grouping process. Categories can be innate, as in the case of the formulation of the existence of universal grammar (Chomsky, 1976). The concept of universal grammar presupposes that languages share an underlying set of rules, that it is further specified in each linguistic system. It presupposes that human beings are born with the ability to detect invariance in the signal and with the ability to use it to generate hierarchically organized sentences, that are possibly infinite, given a finite set of categories and rules, and characterized by the recursiveness.

The categories are innate because of the "poverty of the stimulus," and therefore the categorization process must begin adopting an innate strategy. The majority of categories, nonetheless, are learned through experience and the way we learn them is part of the categorization process, and it will be discussed below. In final instance, categories can be described through 3 properties: discreteness, vagueness, and functionality.

Categories can be discrete and non-discrete. Examples of discrete categories are for example noun and verbs in grammar and phonemes in phonology. They are defined by presenting opposite features across categories and sharing features within category. Therefore, single items are categorically and binarily members or not-members of a specific category. Nonetheless, categories can also be non-discrete, as in the cases of adjectives. Adjectives display features of both the noun and verbs category, so aspects of the verbs can be partially represented within the concept of adjective, and in the same manner, aspects of the noun category also belong to the adjective one. This creates intersection areas that do not allow a binary 1 vs. 0 description. The concept of vagueness expresses the overlap between categories. A member for which it impossible to establish the specific positioning within one single category is vaguely represented and for this reason, is placed at the borders of adjacent categories that gradually merge into one another.

Functionality refers to the property of an item to belong to only one category or more. Homophones often are associated with different grammatical categories and/or meaning. The word "call" in English belongs both to the noun categories "a call" as well as to verbs "to call." In this sense it expresses multifunctionality.

1.3.2. Categorizing

Scientists working in perception, cognition, and philosophy agree with the idea that categorizing is a property of the system or organism that categorizes and not a property of the world that can be categorized (Harnad et al., 2005; Holt & Lotto, 2010; Robert, 2005). The world comes in contact with the interface of the system that can detect it and perceive it, and afterward, during the processing, the system organizes the information it has gathered, in order to store it and make operations with it in a more efficient way. In case of human beings and speech, a continuous complex soundwave comes in contact with the sensorimotor system (E. J. Gibson, 1969; J. J. Gibson, 1979; Harnad et al., 2005; Robert, 2005). Categorization is a process considered at the base of human cognition. The signal processed by the sensorimotor system is mentally reproduced and forms

> "mental entities, which are categories, mental relations between categories, and mental interactions (or operations) on categories. [...] Knowing is categorizing."(Robert, 2005; p.701).

In strictly perceptual terms, categorizing

> "reflects a decision about an object's type or kind requiring generalization across the perceptually discriminable physical variability of a class of objects"(Holt & Lotto, 2010; Palmeri & Gauthier, 2004).

The acquisition of the native language phonological system is based on the innate ability to discriminate a plurality of speech-sounds contrasts present in the spoken languages of the worlds. By the age of 1, the discriminable phonological contrasts are significantly reduced to the contrasts available in the infants' native language (Aslin, Saffran, & Newport, 1998; Jusczyk, 1997; Patricia K. Kuhl, 1991; Polka & Werker, 1994). The mechanism underlying the loss of sensitivity during the course of age is the formation of the categories of the native languages. It is assumed that infants are provided with the ability to extract recurring cues in the incoming signal and through statistical learning can detect which features are distinctive in the language they are exposed to (Patricia K. Kuhl, 2000; Maye, Werker, & Gerken, 2002; Saffran, Aslin, & Newport, 1996). The ability to associate stimuli presenting common features is the ability to "acquire similarity" (Kuhl, 1992) at the basis of category learning. The loss of discrimination sensitivity is then imputable to the use of categories, once they are formed, as a filter in perception, causing the perception of members that prior of the formation of the category were discriminable, and after, are classified as identical by belonging to the category. This perception modality is what is called Categorical Perception (CP) and is discussed in the following section since it is associated with a specific experimental paradigm that is adopted and adapted in this work. A particular aspect that it is worth to mention at this point is that CP assumes that perception is blind within category, implying that there is not a sensitivity within category establishing which members are better members.

A second theoretical approach investigating the role of categories, on the contrary, assumes that categories affect perception but that sensitivity within category remains possible. This is possible because the categories are assumed to generate a sort of template representing the best instance, the prototype, of the stimuli that could be grouped as category members. The prototype functions as a magnet and attracts all the instances resembling it, the instances closer to the magnet are ones resembling it the most. This is what is called the Perceptual Magnet Effect (PME) (Patricia K. Kuhl, 1991).

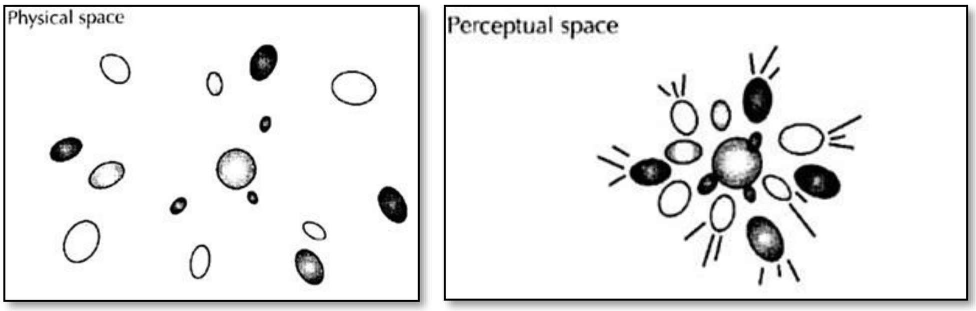


Figure 1.3.1 The warping of the perceptual space - adapted from Kuhl, 1994, p.813

The relevant aspect of both models is that the establishing of stable categories "warps the perceptual space" (Patricia K. Kuhl, 1991)causing stimuli to be categorized differently on the basis of experience.
Language in this sense shapes auditory perception.
Two facts of interest derive from here: why is the loss of sensitivity an advantage? And can it be overcome, and how? Acquiring a new phonological system in a second language implies the establishment of new categories or at least the restructuring of the existing ones. There are available at least two theoretical models that account for the perception modality in L2, the Speech Learning Model (SLM) (James E. Flege, 1995) and the Perceptual Assimilation Model/L2 (PAM/L2) (C. T. Best & Tyler, 2007) that will be discussed in following sections. I will address here first, in a more abstract manner why the loss of sensitivity represents an advantage, in cognitive terms, suggesting it can provide a theoretical departure point also for research in L2 acquisition.
In logical terms, forming a mental representation of the perceived events and reality of the world is defined as an inference. The categorization process can be seen as an inference operation on the perceived stimuli, that associates perception and action(Robert, 2005). The process of grouping several instances within a categorical entity is classified as analogical inference, and it is possible to classify 3 types of analogies (Holyoak & Thagard, 1995; Robert, 2005). An analogy can be established recognizing the exact identity of two instances (Object-sameness) associating them within the same category. A second type of analogy is the one allowing to recognize a relation between categories (Relation-sameness). It is the ability to establish, for example, that what relates a category A to a category B, is somewhat analogous to what relates a possible category C to a category D. In final instance, it is possible to establish a relation between relations (Relation^2-sameness). This third type of analogy is what allows to think causally, so to put into relations causes and effects. Holoyoak & Thagard (1995) and Robert (2005) associate the ability to process Relation- and Relation^2- sameness analogies, with the presence of language. In this logic explanation, categories are filled through generalization processes, inductive arguments. Robert (2005) proposes this example: by encountering several white swans, we can generalize a conclusion that all swans are white. This perceptually reflects a bottom-up statistical analysis of a stimulus, and the categorization process that the category "swan" is filled with animals resembling the features of a swan and that also are "white." The generalization process, nonetheless, generates a second implication, because by creating the category "swan" with the property of "being white," we also predict that all the swans, that we could encounter in the future, will be white. This reflects a top-down activation that generates an expectation attached to the category swan to resemble the prototype of the category. What happens if we encounter a swan that is black? The generalized concept is corrected and reduced to "most swans are white." The process of correcting an inference causes that the previous chain, connecting only two points: swans and the property of being

white, now it is split in 3 links, the category swan and the possibility to be both black and white. This mechanism, called abductive correction, is for Robert (2005) the mechanism at the base of cognition:

> "Cognitive progress results from changes in our system of categorization that increase its explanatory and predictive power. This progress takes the form of a refinement of our system of categories and their causal relations." (Roberts, 2005; p.714)

According to Robert (2005), we use this mechanism to compensate for reduced memory capacities, since it allows to construct a concept, simpler to store when compared with a plurality of unorganized entities, that generates the possibility to access the members of that categories in an organized way. In second instance, it allows comparing larger groups of organized entities, that now are possible to be distinguished by the process of "acquired distinctiveness." By constructing a hierarchical organization of categories. Therefore, we can reduce the quantity of information and remember only the salient aspects necessary to access the cognition. What remains unclassified is treated, thus, as an irrelevant detail, that can be forgotten.

> "knowing starts by forgetting most details and using only the most salient aspects of our sensations to construct categories" (Roberts, 2005; p. 715)

The elegant logical model just described patterns with factors playing a role in linguistic auditive perception. The mechanism accounts for the formation of a stimuli-driven category, and for the activation of the category once it is established generating predictions on the properties of its member. The idea that the hierarchical organization of categories allows making use of salient aspects efficiently is what is commonly described in the field of attention.

The mechanism that selects a portion of information in the signal is

> "the combination of a rapid bottom-up saliency driven (task-independent) attention, as well as slower top-down cognitive (task-dependent) attention (Itti, Koch, & Niebur, 1998).[...] Similarly, in audition, one may hear people talking, music playing in a room (saliency-driven), but it will not be immediately apparent what people are saying or what type of instruments are producing the music. Only if the subject chooses to listen to the music, s/he will be aware of what kinds of instruments are producing the music (top-down)." (Kalinli & Narayanan, 2007; p.1).

In first instance, the stimulus-driven bottom-up processing occurs, and it is guided by attention towards salient features in an unconscious manner; secondly, the top-down process voluntarily shifts the attention toward features considered relevant. The mechanism causes that only the attended features can progress to high-level processing(Alain & Arnott, 2000; Harding, Cooke, & König, 2007; Itti et al., 1998; Kalinli & Narayanan, 2007).

Categorization, therefore, is a mechanism needed to avoid the processing of irrelevant information. We form categories in order to organize the information and use the organization of knowledge that we have constructed to select relevant information in the incoming stimuli. We can correct and update the organization of information we have formed, by noticing incongruencies between the predictions that are established by the category and the perception of the incoming stimulus.

The native language phonology, therefore, functions as an auditive stimuli "organizer" and we use it to identify recognizable linguistic input in the signal – categorical perception – and we must update the criterion of saliency in order to form and establish new categories that are necessary to create the phonological contrast in a second language.

1.3.3. The Categorical Perception Paradigm

The notion of CP emerges from the research at the Haskin Laboratories in the 1960s. Liberman et al. (1957) constructed a series of stimuli built with syllables composed of a vowel approximating the vowel /e/ and an onset consonant that could span from /b/ to /d/ and /g/. The consonant shift was obtained by continuously increasing the onset frequency of F2, the second formant. They found that participants classified them in the 3 categories corresponding to the 3 existing consonants in English.
In addition, the authors tested the discrimination sensitivity of participants through an ABX task, that requires to judge the third stimulus of a series as equal to stimulus A or B, previously presented. They found that participants, when presented with stimuli belonging to different categories, could discriminate them easily, on the contrary, the stimuli belonging to the same category, despite the same physical difference, were not accurately discriminated. This paradigm has been used in a conspicuous number of studies replicating the effect and therefore it has become a standard paradigm in the study of auditory perception (Abramson & Lisker, 1970; Bastian, Eimas, & Liberman, 1961; Alvin M. Liberman, Harris, Kinney, & Lane, 1961; Mattingly, Liberman, Syrdal, & Halwes, 1971). In experimental terms, the paradigm is composed of two tasks: a categorization/identification task and a discrimination task. In the categorization, task participants create a distribution of nominal answers, by assigning a category to a repetitive battery of stimuli. For each category, the proportion of answers of each participant can be fitted with a logistic, sigmoid function. In the shifting continuum of stimuli, the stimuli receiving 100% of answers, identifying category A, e.g., the consonant /b/, are represented with left bottom tail of the curve. Stimuli 100% identified as category B, e.g., the consonant /d/, are visually represented in the top right tail of the curve. The stimuli placed in the middle of the continuum (in terms of PME, bad exemplars for both the category and therefore lying at the categories' boundaries) are partially identified as category A and partially as category B, determining the turning point of the curve.
The curve is then identified by two parameters: the position of the turning point (α parameter) and the slope of the curve (β parameter).
The discrimination task, on the other hand, generates a function of correct identification, or accuracy in the discrimination. The accuracy when classical CP occurs then displays to be low, or at chance, for stimuli classified as A, and for the stimuli undoubtedly classified as B, and increases abruptly in the position of the turning point, creating a peak.
CP Perception, thus, is traditionally assumed to be attested by the presence of 3 hallmarks within the two tasks: (Holt & Lotto, 2010):

1) Listeners' responses in the categorization task fit the shape of a sigmoid function, abruptly shifting between the two (or more) experimentally available categories (e.g., phonemes) in the range of a small acoustic change, defining the perceptual boundaries of the two categories;

2) Listeners’ responses in the discrimination task fit a discontinuous function displaying a peak of correct hits when two stimuli lying on the opposite sides of the boundaries are presented, and a low rate of correct responses (flat contour in the response function) when stimuli within category’ boundaries are presented;

3) Categorization results predict discrimination results, establishing the dependency of listeners' discrimination performances to the listeners' existing perceptual categories in that specific perceptual space.

The original definition of CP from the Haskin group is reported below:

> "Categorical perception refers to a mode by which stimuli are responded to, and can only be responded to, in absolute terms. Successive stimuli drawn from a physical continuum are not perceived as forming a continuum but as members of discrete categories. They are identified absolutely, that is, independently of the context in which they occur. Subjects asked to discriminate between pairs of such "categorical" stimuli are able to discriminate between stimuli drawn from different categories, but not between stimuli drawn from the same category. In other words, discrimination is limited by identification: subjects can only discriminate between stimuli that they identify differently." (Studdert-Kennedy, Liberman, Harris, & Cooper, 1970; p.234; in Repp, 1984; p. 252)

The presence of CP, together with another important phenomenon pertaining segments – coarticulation - has led to the formulation of the Motor Theory of Speech Perception (MTSP) (Liberman et al., 1967; Alvin M. Liberman & Mattingly, 1985). The MTSP assumes that

> "When acoustic patterns are different, but the articulatory gestures that would have caused them in natural speech are the same, or vice versa, perception tracks articulation (Liberman, 1957)" (Galantucci, Fowler, & Turvey, 2006, p.362).

Therefore, according to the MTSP what grants CP are the articulatory gestures that generate the sounds. It followed that according to the MTSP, CP it has been regarded as a prerogative of speech only because only speech sounds are predicted to be mapped in articulatory movements of the vocal tract. The MTSP, moreover, developed in the years of a paradigmatic shift in psychology, from the behaviorist to cognitivist approach, and therefore adopted also a modular view of cognition. This implied that speech perception could be performed given the presence of a phonetic module responsible for the decoding of sounds in language and CP had to be mapped within this theoretical framework and therefore be classified as something related to language.

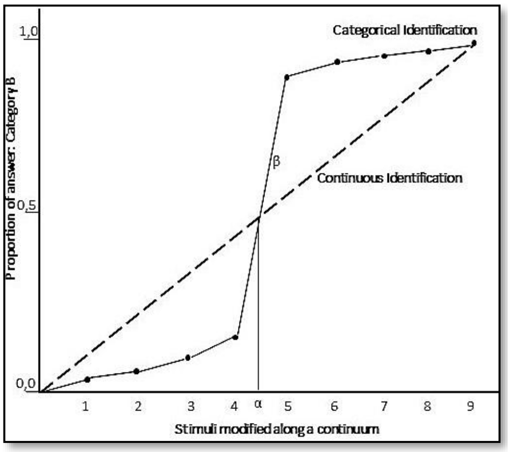
The MTSP received several critics along the years, due to the evolution of the psychological science and due to the advance of technology and the study of animal cognition. Nonetheless, Liberman and Mattingly, the authors founders of the theory, have the merit to have always updated the theory, in order to respond to critics (for an extensive review see (Galantucci et al., 2006)and Liberman (1996)), and for what pertains the perception of consonants the theory has kept a high predictive value.

For what pertains the activation of motor areas during speech perception, this has been found valid, nonetheless the current neuroscientific models of auditory perception, do not support the view that the motor representation is a necessary mediation during auditory perception, but instead propose that two neural networks process sound: the dorsal and the ventral stream, supporting different functions (Hickok & Poeppel, 2007). According to the dual-stream model the mapping between acoustical features of segments, and they motoric representation of the vocal tract movements is processed by the dorsal stream. It is not, nonetheless, intended that the dorsal stream acts as primary auditory area specialized only for speech.

What it is interesting for this work, that emerges from the early studies of speech perception, is the idea that language phonetic features sort of "have to" display signs of to be categorically perceived, and despite the presence of criticisms and the findings of controversial results, this notion somewhat resisted in the literature, so that one of the first checks in order to establish a linguistic function of a phenomenon, is the test for the presence of CP. This has also happened for intonation, finding a non-clearer picture and creating a debate on its nature.

1.3.4. Continuous and Quasi-Categorical Perception

Perception that does not display the discontinuous modality granted by the presence of the categories is considered Continuous Perception (CPe). If the two tasks composing the CP paradigm are performed, CPe emerges if the proportion of answers identifying a specific category gradually increases from 0 to 1 (0% - 100%) while the presented stimuli shift from category A to category B.
In addition, the discrimination sensitivity shows not to be related to the presence of a category, so it is independently and constantly high or low.
Continuous perception is known to characterize phonetic aspects such as loudness, or duration when it does not represent a distinctive feature in a specific language (Bastian & Abramson, 1962).
The controversies pertaining the notion of CP emerged because not all the linguistic features appeared to be categorically perceived.
For example, the identification of vowels has been found to shift along the continuum of stimuli gradually, and their discrimination has been found equally high (Eimas, 1963; Fry, Abramson, Eimas, & Liberman, 1962; David B. Pisoni, 1973).
Specifically, pertaining pitch, in this case with lexical function: it displayed CP when stimuli ranged from a level tone to a rising contour tone, but non-CP, when stimuli ranged from a level tone to a different level tone (Abramson, 1978; Francis, Ciocca, & Ng, 2003; W. S.-Y. Wang, 1976);



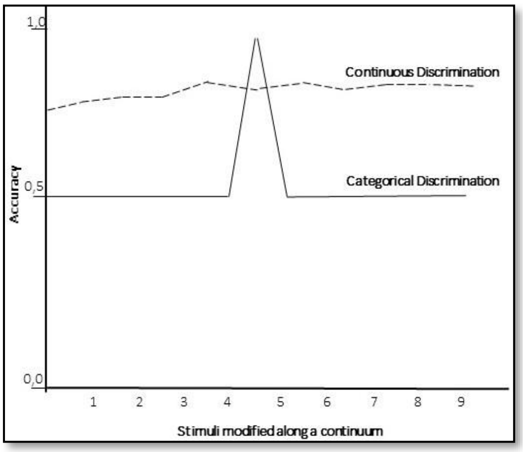


Figure 1.3.2 Categorical and Continuous Identification and Discrimination

The perception of pitch modulation at sentence-level has been tested in several experiments that are presented in the introduction of the first study (Chapter 2). The debate on the nature of intonation is generated by the proposal that intonation could be either categorically, continuously, or quasi-categorically perceived.

Perception can also display properties of both categorical and continuous perception. This is the case when the identification task is associated with a pattern of answers that can be fitted with a logistic curve, but in the discrimination task, the accuracy is neither completely independent from the category nor significantly taking advantage of the category shift.

Instead, it happens that the performance of the participants gradually increases in the perceptual space that still is classified as category A, keeps a high rate of accurate discrimination across the category boundary, and decreases within the perceptual space identified as category B.
The discrimination function in this case, instead of forming a peak associated with the turning point is characterized by a graph of the shape of a plateau.
The implications of quasi-categorical perception are unclear but probably are best explained with the PME model. If a representation of a prototypical member of the category exists, it is possible as the perceptual distance from the prototype exists, thus, when shifting towards the category boundary, the discrimination sensitivity can take advantage of the increasing distinctiveness. The discrimination accuracy reaches its maximum and then gradually decrease when approaching the prototype of the category B.
This implies that participants can disentangle the phonetic details and are able to compare them with the ideal prototype of the phonological representation.

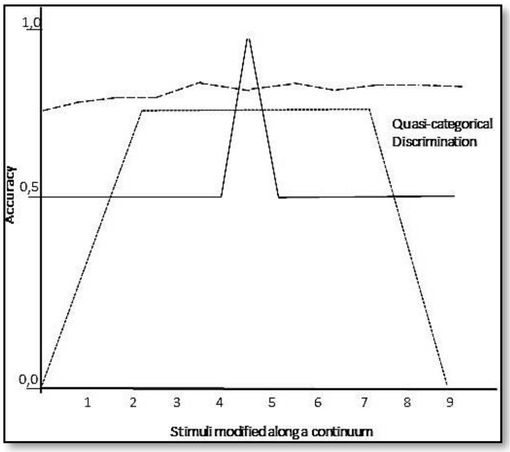


Figure 1.3.3 Quasi-Categorical discrimination

1.3.5. The perception of prosodic categories: the special role of alignment

Prosodic categories entail a higher number of dimensions when compared to segments. They contemporarily display a "holistic, gestalt-like," (Schneider, 2012; p.11) aspect, but at the same time, they can be decomposable, as described in the introductory section on AM, in discrete entities with a phonological identity and carrying a morphemic value. (Von Heusinger, 1999). What allows to switch from pure acoustic processing to a linguistic one is the complex mapping of tune-text association. This implies that the categorization process has to map, and be mapped, considering both the phonetic aspects of the pitch contour, syntactical and metrical structure allowing an abstract representation of the accents, and an integration with the semantic content of the utterance, allowing the integration of the full discourse model, the resolution of anaphoric and cataphoric inferences, and the evaluation of the speaker's deliberate intention to highlight specific syllables within specific prosodic units, and therefore specific referents and meaning.
Alone within the native language system, this implies an incredibly higher number of dimensions guiding pitch modulation, and therefore the decoding of possible interpretations; it implies that these dimensions have to be evaluated within category (e.g., is the perceived pitch stably high or stably low?) but also that the phonetic features of each category must be compared with referential pitch values of the highest

prosodic unit (e.g., is the perceived pitch stably high in a target position set by metrical and syntactical organization? Stably higher than average reference value considering the natural declination effect? Stably higher than a reference value congruently with the content of the utterance? Stably higher than a reference value congruently with the intentions of the speaker?)

This is an aspect of perception addressed by the notion of intrinsic and extrinsic reference (Cutting & Rosner, 1974; Miller, Wier, Pastore, Kelly, & Dooling, 1976; D. B. Pisoni, 1977; Xu, Gandour, & Francis, 2006).

An intrinsic reference is a comparison between two levels of an acoustic parameter switching across time (e.g., a formant transition, the temporal order of two acoustic cues). The variation in intrinsic reference perception has to be detected evaluating one stimulus at the time, demanding a low cognitive load of working memory. Perception of events variating in intrinsic reference modality has been generally found as categorical (Cutting & Rosner, 1974; Miller et al. 1976; Pisoni, 1977).

F0 modulation alone (absence of segmental information) is considered a parameter processed through intrinsic reference (Xu et al., 2006) and therefore the categorical perception of non-speech stimuli variating within two pitch levels is considered favoring Categorical Perception, despite being outside the domain of speech. For example, CP has been found for musical intervals (Krumhansl, 1991) and simulated sounds of music instruments (Cutting, 1982; Cutting & Rosner, 1974), artificial stimuli (Livingston, Andrews, & Harnad, 1998), and non-speech continua (Xu et. al 2006, Miller, 1976).

On the contrary, the perception of pitch in the presence of segments displayed variation on the basis of the nature and function of pitch modulation. CP has been found for example in the perception of tones with lexical functions in tonal languages. In particular, CP has been found for stimuli that ranged from a level tone to a rising contour tone, but on the contrary, non-CP was found when stimuli ranged from a level tone to a different level tone(Abramson, 1978; Francis et al., 2003; W. S.-Y. Wang, Lehiste, Chuang, & Darnovsky, 1976).

The specific findings on the perception of final BTs at sentence-level will be addressed in the next chapter. Nonetheless, it is possible to affirm that the main issue concerning the perception of F0 movements at sentence level is the lack of consistent findings in either one of the directions: categorical or non-categorical, which led to the postulation of the concept of quasi-categorical perception described before.

A possible explanation for quasi-categorical perception is the memory load necessary to classify a pitch movement in a discrete way. The presence of segments and the evaluation of pitch height having as reference the prosodic unit demand, in fact, a process of extrinsic reference.

Extrinsic reference refers to the evaluation of specific auditive levels of a parameter that require multiple trials in order to be classified. They require the capacity to keep previous trials in memory and to evaluate more than one dimension that can vary within stimulus over time. This is the case for example of steady-state non-speech continua, the formant contrast of static vowels and pitch comparisons of level tones inhibiting intrinsic reference (Mirman, Holt, & McClelland, 2004; Xu, Gandour, & Francis, 2006).

A second phenomenon impacting on the perception of F0 is the presence of a Critical Bandwidth (CB) and of resolved and unresolved harmonics (Fletcher, 1940; House, 1990). The CB corresponds to the minimal invariant distance within the cochlear membrane necessary to discriminate two frequencies values. As frequency values increase the available surface of the cochlear membrane decreases, causing high frequencies to overlap in position falling within the CB and decreasing, therefore, the discrimination capability.

> "If two sinusoids are presented to a listener at the same time, the listener will hear two separate pitches if the frequencies of the sinusoids are further apart than a critical bandwidth. If the

sinusoids are not separated by a critical bandwidth, the listener will no longer hear two separate pitches." (House, 1990; p.24)

The structure of the cochlea, therefore, causes that the CB is narrower for lower frequencies compared to higher ones.

In speech, lower harmonics benefit from this phenomenon granting them to be sufficiently far apart to be distinguishable; on the contrary higher harmonics fall within the CB. Lower harmonics are then classified as *resolved harmonics* and upper ones as *unresolved harmonics* (House, 1990). The presence of segmental information, therefore, creates a more complex stimulus in comparison to a stimulus where only F0 modulation is present because of the presence of high unresolved harmonics.

> "The existence of high-order unresolved harmonics in the speech stimuli reduces the ratio of spectral energy distribution in low-order resolved harmonics. Given the equivalent overall spectral energy between speech and nonspeech, it yields a lower pitch salience for the speech stimuli because resolved harmonics contribute more to pitch perception than do unresolved harmonics(Shackleton & Carlyon, 1994; Stagray, Downs, & Sommers, 1992)" (Xu et al., 2006; p.1072)

Further on, in a sequence of studies belonging to his Ph.D. dissertation, House (1990) tested the sensitivity to pitch movements aligned to vowels, under a specific hypothesis: The Spectral Constraint Hypothesis (SCH).

The starting assumption is that speech, considered in a schematic way a sequence of consonants and vowels (CVC), the spectral information considerably varies through time in the passage from consonants to vowels. The spectral information is determined by the presence of several biomechanical factors required by the act of speaking. Speech is characterized by the presence of aspiration, increasing intensity, and the mechanisms underlying the production of plosives, for example, that are characterized by a phase in which air is blocked and then abruptly released. Therefore, in the temporal phases between consonants and vowel, there is a rapid change of formants transitions representing the maximum spectral information, detectable at the onset of the vowel. (House, 1990). This determines that if pitch is variated at the beginning of the vowel perceptual changes occurs on two dimensions: at the pitch level and at the spectral information level.

House (1990) experiments found that the perception of pitch movements (rise-fall and fall-rise configurations) aligned to only one steady vowel /a/ can be categorically identified as movements, and the two patterns are correctly discriminated from one another. By introducing spectral variation (CVC configuration), the discrimination sensitivity towards pitch decreased.

When pitch movements were aligned to the onset of the vowel (/abaa/), some participants began to invert the classification of the rise-fall, fall-rise patterns, some were able to adopt the same strategy as in the first task, some were no longer able to classify the stimuli in discrete categories.

When pitch movements were aligned to the end of the vowel of a CVC configuration (/aaba/), the end frequency value of the pitch movement aligning to the end of the vowels assumed a reference value for the categorization process. So again, the results could be divided into three types of performances: those who were able to categorize the movements as in the first task, those categorized on the basis of the end frequency value, and those who were not able to categorize.

When pitch movements were aligned to the center of the vowel, within a CVC configuration (/aba/), the results were similar to the ones obtained with the alignment at the end of the vowel, therefore, favoring the end frequency values as target referent.

House (1990) concludes that in the absence of spectral change the perception of tonal variation is enhanced, and he postulates that the mechanism can be relevant for speech since it is possible to assign a distinctive value to pitch movements aligned to vowels, as it is in tonal languages for example.
Nonetheless, he highlights that individual variability is highly present in his sample of participants since in truth one-third of the perceivers remained able to correctly identify pitch categories, based on pitch modulation, also when aligned to high spectral variation.
The SCH thus postulates that:

> "areas of the maximum new spectral information influence and constrain the perception of tonal movement. [...]. Tonal movement through spectral change is recoded as pitch levels, while movement through spectral stability is coded as movement configuration." (House, 1990; p.81)

In the second sequence of experiments, House (1990) investigated the role of pitch modulation on phrasing. He created stimuli where the word 'five' was repeated 5 times, and different pitch modulations were aligned on each word. The task consisted in asking participants to group the 5 repetitions of the word 'five' in a 3+2 configuration or, 2+3 configuration.
He aligned to segmental material five types of contours: rise-fall, fall-rise, falling, rising, steady modulations. The results of House (1990) highlight that steady configurations, level tones, were preferred as cues for phrasing when compared to movements. Also, that rapid fall-rise and rise-fall movements aligned to one single syllable are economically reconfigured as level tones, where on the contrary, slow unidirectional rising or falling movements are interpreted, as actual pitch movements.
From derives the Tonal Movement Coding Hypothesis (TMCH) stating that:

> "a relatively rapid tonal movement at the beginning of each element [...] is recoded in terms of a relative pitch level, a while slower movement [...] is coded as a movement configuration." (House, 1990; p.100)

In House's results, nonetheless, is also present a deviation from this starting hypothesis. Namely, that

> "a movement configuration can be recoded also in terms of relative pitch levels with the tonal fall functioning as probable low and the tonal rise as a probable high." (House, 1990; p.100)

In conclusion, the results highlight that a movement, occurring at a slower rate (slowly rising or slowly falling), is most probably perceived as movement. Thus, in a congruent configuration with the first hypothesis the presence of a vowel, given its spectral information and the possibility to stretch in duration allows for the highest sensitivity of pitch movements.
Addressing in final instance the topic of pitch perception in speech, House (1990) postulates the presence of a Precategorical Acoustic Storage memory (Crowder & Morton, 1969) ,supporting the presence of an early 'frequency sweep detector' (Taylor & Wales, 1987; Wales & Taylor, 1987), that has the function to retain pre-analyzed frequency movements that can be then integrated once the segmental material is processed.
In the presence of a short segment, short syllable or short vowel within the syllable, the system economically stores rapid movements as levels, in the presence of sufficient duration and spectral stability pitch modulations can be stored as movements.
For House (1990), this is the basic process at the basis of categorical perception of intonational categories.

Interesting evidence on the perception of stimuli varying on a multidimensional space, contrasting infancy and adulthood perception comes from Goudbeek, Smits, Cutler, & Swingley (2005).
They adopted the conceptualization that the evaluation of a sound corresponds to the mapping of a point on a multidimensional psychophysical space. The repeated exposure to several sounds forms a perceptual "cloud" of points in the psychophysical space. The categorizing process, in these terms, is the association of a perceptual cloud to a category after the exposure to several trials (Behnke, 1998; Kornai, 1998).
They compared the perception of stimuli constructed with a unidimensional variation (clouds displaying variance only in one direction, caused by the variation of one single parameter) with stimuli generated by the combination of two dimensions (clouds varying in functions of two dimensions, therefore graphically displaying variance on diagonal vector) in adult perceivers. The starting hypothesis, deriving from the critical period conceptualization is that adults lose along with age a detailed sensitivity and ability to categorize sounds shifting on a multidimensional space, where on the other hands, this appears to be an innate ability in infants.
The hypothesis then was that stimuli constructed on two-dimensional perceptual space would be categorized with lower accuracy.
The investigated the phenomenon, firstly with non-speech material, with training sessions granting feedback of correct or incorrect categorization. In the first condition, stimuli varied only along one dimension, duration; in the second condition, they varied along two dimensions: the frequency of the spectral peak and duration.
The task was to categorize the stimuli adopting two levels, A and B, and it was performed in the two conditions.
The results highlighted that through the training and in the test phase, participants were able to learn and accurately categorized stimuli shifting only along one dimension. On the contrary, in condition 2 no-learning effect is detected, and the test phase highly resembles the poor performance of the training session.
They performed the same task adopting speech material. The stimuli were constructed adopting vowels in Dutch modulated by the same two dimensions adopted in the non-speech task. The results highlighted the same pattern emerged in the previous task; thus, the categorization process displayed a pattern of congruent acquisition only when the classification pertained the aspect of vowel duration, and accuracy significantly dropped when both duration and type of vowel changed.
These results seem to sustain the hypothesis that the ability to categorize newly perceived stimuli on the basis of a perceptual intersection between existing categories strongly decreases with age (even though in this experiment infants are not compared to adults to demonstrate a significant difference).

1.3.6. A cognitive interpretation of Categorical Perception of Alignment

Parallelly, to the AM theoretical models, research has developed investigating the properties of *tonal alignment*. (House, 1996, 1999; Kohler, 1987; Niebuhr, 2003, 2006, 2007).
In one of the first studies, Kohler (1987) investigated the perception of utterance "sie hat ja gelogen" ('she has been lying"). The sentence is realized with the stress on the syllable "-lo," and according to the theoretical model adopted by Kohler it can be functionally interpreted in 3 ways according to 3 alignment positions: with an early peak actually positioned in the preceding syllable, with a medial peak centered on the target syllable, and with a late peak at the end of the syllable. According to the model adopted by Kohler, the 3 different functional interpretations respectively are: "an established fact," "new information," and "an emphatic contour." In the experiment, 11 levels of stimuli were created shifting from early to late peak. In a first experiment, participants heard prior to the exposure to the first 8 levels (ranging from early

to medial peak, and from established fact to new information) , a context sentence: the utterance "jetzt verstehe ich das erst" (now I understand), which guided the possible interpretations of the target sentence preferably towards either the interpretation as new information. Participants after each trial had to state whether the intonational contour that they heard, matched the pragmatic context. In a second experiment, he performed a discrimination task, pairing stimuli distant from 0 levels to 2 levels. Participants had to state whether they had perceived a difference in the pair of stimuli. The results of this first experiment reported an effect of CP of alignment. The identification task displayed a sigmoid curve for the identification of the correct accents for the context, and the rate of accuracy in discrimination in the transitional area between categories was significantly higher than within category.

In a following study, previously mentioned in section 1.2.7. discussing the concept of plasticity, D'Imperio & House (1997) investigated the perception of alignment in Neapolitan Italian focusing on the disambiguation of statements vs. questions. In Neapolitan Italian, broad focus sentences are produced with a falling pitch contour (H+L*) and narrow focus sentences with a high peak H* (D'Imperio, 1995). This can also be interpreted as a difference in alignment since the peak that in broad focus sentences is aligned earlier, causing that the falling contour crosses the accented syllable, actually shifts in the middle of the vowel reaching a peak followed by an immediate fall on the target vowel. In Neapolitan Italian, a similar phenomenon occurs in the realization of Yes/No questions, where a rising contour is aligned in nuclear position (L+H*), and it is followed by a fall, generating overall a contour in which the typical questions' rise is absent (D'Imperio, 1996). On the basis of House (1990)'s results, the hypothesis that the information for a categorical shift from statements to questions' interpretation is carried by the modulation of the rise within the tonic syllable. In this case there were two hypotheses leading the study: one was the alignment position of the L+H* peak, which can be seen as the H* peak of narrow focus sentence realized even later in the syllable, the second one presupposed the mere presence of a rise in nuclear position could gran the identification of a question. The authors recorded two prototypical affirmative and interrogative sentences in Neapolitan Italian and manipulated the stimuli generating 20 tokes. The affirmative prototype was manipulated shifting first, only the peak in 5 positions over time, and the shifting the anchoring point at the end of fall in 5 positions, creating 5 tokens falling with 5 different slopes. The same procedure was done for the interrogative stimulus. The stimuli all contained the sentence "Mamma andava a ballare da Lalla" ("Mom used to go dancing at Lalla's"). The stimuli in this way are composed by a peak sharing the same shape in all tokens that shifts in position along the accented vowel in "lal-la." The early alignment of the peak should grant the perception of a statement since the vowel is interested by a falling contour, the late alignment of the peak should grant the perception of a question since a rising contour crosses the vowel. After each stimulus participants were asked to perform a classical identification task and to state whether it resembled a statement or a question.

The result highlighted the predicted effect, and the category statement was 100% identified in the early alignment condition, and approximately never identified in cases of late peaks. The stimuli with which participants showed higher uncertainty were the ones in which the peak was centrally aligned in the vowel and therefore were connotated by both a preliminary rise and a following fall with the target syllable.

The interpretation of the results is furnished adopting an explanation in line with the concept of extrinsic reference comparison. Namely, that perceptually, the pragmatic interpretation of a tune is accessed through the consideration of the overall contour, in a way that it is the synergic positioning of both low and high tones that grants a categorical interpretation to the presence of a peak. On the basis of these results, the label H* for narrow focus sentences is proposed to be substituted with H*+L to signal the presence of a final low target point.

Niebuhr (2003) adopting the paradigm proposed by Kohler (1987) and House (1990) investigated the perception of 4 pitch contours differing in alignment and phonetic realization of the peak shape in German. He adopted a stimulus like *"Sie war mal Malerin"* (*"She was once a painter"*) and the following 4 shapes: fast rise and slow fall (f/s); slow rise and fast fall (s/f); fast rise and fast fall (f/f); and slow rise and slow fall (s/s). The 4 types of contour then where resynthesized in 11 levels shifting the anchoring position of the peak, and ranging from early, to medial, and late peaks. The stimuli were presented to participants with a preceding context sentence as in Kohler (1987) creating an expectation for new information in the target sentence. Participants had to judge whether the target sentence correctly matched the context set by the prior sentence.
As in the paper of Kohler (1987), the early alignment of a peak is assumed to convey information considered settled and not changeable and the presence of a medial peak, new information. So, the identification task perceptually required to identify a medial peak and in addition, a classical discrimination task was performed. The results highlighted an interaction of both alignment and peak's shapes factor on the categorization process, with alignment having a stronger influence. In the first half of the alignment continuum, slowly falling contours showed a higher discrimination rate. The presence of a slow rise aligned early after the vowel onset reduced the identification of medial peaks and the stimuli composed by slow rises and slow falls where judged with higher uncertainty and disagreement by perceivers. The interaction emerged in the perceptual positioning of the categories' boundaries since it appeared that the presence of a fast rise, combined with a slow fall, shifts the category boundary towards the category "early peak" increasing the perceptual space in which the stimuli are identified as "medial peaks." Thus, a fast rise followed by a slow fall induces more the perception of a medial peak. The opposite effect is found for contours having slow rises and fast falls that shift the category boundary toward the right end of the graphic representation of the sigmoid pattern of the answer, therefore inducing the identification of a wider range of stimuli as early peaks. Pointed peaks composed of fast rises and fast falls are recognized with a higher interindividual variability that slow rising and slow falling peaks.
The categorical perception of the 3 positions of the peaks: early, medial, late, contributed to the grounding of the Kiel Intonation Model (KIM) (Kohler, 1991, 1997). In the KIM model, the 3 alignment positions have a phonological status.
The same paradigm has been adopted in two following experiments to investigate whether the alignment of a valley (a low tone anchor point) showed categorical properties in function of the early, medial and late positioning of the L* target. Landgraf (2003) and Redi (2003) independently found that alignment in case of the presence of a valley is not categorically perceived, and instead are characterized by continuous perception.
Niebuhr & Kohler (2004) investigated the interaction of a context sentence modulating the position of a valley and inducing an expectation in the pragmatic expectation of the content that follows in the target sentence, again discriminable in this case only on the basis of the identification of a valley. The identification task required to state whether the target sentence matched the context induced by the prior sentence, the discrimination task required to state whether the pair of stimuli were identical or different on the basis of the aligning position of the valley. The results highlighted continuous perception again. The stimuli were correctly discriminated with a high percentage of accuracy. The identification function did not display an abrupt transition from non-matching to matching categories, but the presence of the two categories is observable.
The explanation provided by Niebuhr and Kohler (2004) for the absence of a clear-cut CP, considers the mapping of intonational contours with the pragmatic interpretation of the sentence, defined as a

perceptuo-cognitive task. A task of this type requires for the authors the evaluation of each contour with a prototype contour associated with the pragmatic function induced by the context.
A sharp distinction between two categories generated on the presence of a valley is impaired by the following circumstances (Niebuhr and Kohler, 2004; last page of a proceeding):

1. "The perception of early and late valleys depends on more than one salient cue (position and duration);
2. Some of these cues are not points in time but periods of time;
3. The decisive differentiator between early and late is linked inside the vowel which lacks local landmarks; so, discrimination is not supported by a prosodic-segmental link at a certain position in the acoustic continuum and is consequently similar and low all the way through the series: an acoustic continuum is mapped onto a perceptual continuum.

> "[…] This more differentiated and cognition-orientated treatment of categorical perception challenges the traditional view of phonology as mediator between acoustic continua and categorical perception: scalar perceptual changes do not preclude the existence of underlying phonological categories and vice versa. Thus, the categorical perception concept needs revision." Niebuhr and Kohler, 2004; last page of a proceeding).

1.3.7. On Phonetic Details and Exemplar Memory

A theoretical position that approaches the processing of speech from a different perspective from the one adopted as a starting point here is the Firthian Prosodic Analysis (FPA) (Hawkins, 2003; Hawkins & Smith, 2001). FPA aims to consider as guiding theoretical framework, not the presence of two distinct domains of analysis, such as the phonological and the phonetic one, but rather placing the phonetic details present in the signal, at the base of the model. The reasons to propose a model of speech perception based on this assumption are somewhat related to the topic addressed here on the nature of intonational categories. Nonetheless, this specific position aims to provide a framework for speech perception overall, thus, focusing on both phonetic details present in segmental information and suprasegmental features.

In this work, I wish to address the issues related only to intonational categories; therefore, I will not discuss the implications of this perspective with the aim of supporting or disagreeing with the whole scope of this theoretical framework. Since this model aims to address the overall processing of speech, it takes a sharp position, questioning, for example, the need for abstract representations, such as phonemes, to mediate the process from sound to meaning.

> "phonemes are not represented: they cannot be, because phonemes are by definition devoid of current hierarchical context: their relational attributes are part of a different contrastive system." (Hawkins, 2003; p.387)

I cannot adequately address here a model, with its implications, that considers both segmental and suprasegmental aspects. Nonetheless, this perspective reaches a set of conclusions that are in line with the line of argumentation proposed here for prosody, that are worth to mention here.

The grounding concept of this theoretical perspective is that there has to be a way to map phonetic information and meaning directly. In this sense, the model assumes the perspective to consider the interaction of the semantic access with the sound domain, that is generally accepted for prosody, because of its sentence-level domain, but that is rather debated for the processing of segments, especially in a domain lower than the word. An example provided by Hawkins (2003) is the different phonetic realization of the segments in a sentence like "I do not know" that can be uttered, for instance, as "dunno." The choice of the phonetic realization of the segments in this utterance, according to by Hawkins (2003) and Hawkins & Smith, (2001), is guided by the presence of contextual information. Thus, in line with the argumentation that I am making for intonational categories, FPA postulates the role of pragmatics for the entire mechanism of speech perception.

The processing of the context of use enters the model through the presence of what is called episodic or Exemplar Memory (Goldinger & Azuma, 2003; Nosofsky, 1991). Exemplar memory stores a frequency distribution of the experienced sound realizations, and thus the adoption of the exemplar memory in a speech processing model allows accounting for the storage of voice quality, pitch range, rhythmic aspects and idiomatic phrases of an interlocutor and for its consideration during the processing of speech. According to the FPA perspective then, through the memorization of phonetic details, we are able to develop fine-detailed predictions on both linguistic and paralinguistic information and to associate them to specific speakers and contexts of use.

Similarly, to the line of argumentation that will consider here the sound domain in conjunction with the discourse structure domain, for intonational categories, the FPA postulates the need for seeking a perceptual coherence in the signal during its processing. There has to be, thus, the co-occurrence of several phonetic features that coherently map with the exemplars stored in memory (e.g., the spectral information coherently has to reflect the biomechanical features of the vocal tract of one single person, the one presenting the stored quality of voice).

This model, thus, postulates, that the nature of speech categories is bound to its context of use and that categories exist only by virtue of it. (Hawkins, 2003).

Drawing upon these concepts, the FPA is provided with a hierarchical structural organization that has to accounts for the systematic relationships in language, as I have previously addressed here (section 1.1., 1.2 and 1.3). The hierarchical structure accounting for the sound domain and grammatical aspects of a language assumes then a Polysystemic (Polysp) nature. The sound domain and the grammatical structure are subdivided into several subsystems having a restricted formal linguistic domain (e.g., function words vs. content words). Each subsystem can express its features through specific phonetic details, and each detail can provide information for more than one subdomain. In this respect, this model recalls what has been previously described in section 1.1 and 1.2. for the structure of the intonational system in language. The representation of a whole utterance is expressed, then, by the set of connections between the phonetic details and the subsystems to which they link, trough "tree-structures." The constructed hierarchical structure is formed by nodes, similar to the ones proposed within the AM framework (e.g., syllable, foot, and IP) but that reaches the segmental properties too, such as the onset, nucleus, and coda, within the syllabic structure. Within this representation the phonemes are not represented.

Polysp is, thus, a system that considers phonetic details the information on which to construct cognitive categories of speech. Maintaining the level of the phonetic information overcomes, for instance, the issues related to the role of alignment within a theoretical model of perception. Since the category is formed by coherently mapping several levels of phonetic information with the possible contexts of use, the time domain is coded within the category itself, reflecting the point of view expressed by Niebuhr and Kohler (2004) on the debate of CP of intonation.

On the other side, I believe it requires higher computational power, compared to the AM approach discussed in this thesis, since a category in this way is defined by the calculation of all the combinatorial configurations of the phonetic details with each context of use. Within an acquisitional perspective, as it will be discussed in the section 1.5., the acquisition of the speech categories of a foreign language, adopting this perspective, must entail the information about the stored coherent structures of the speech categories in the native language. This means, that aside from what it is classically considered the role of phones and phonemes in a generative framework, within this perspective a speaker stores also the properties of the voices of each speaker, in each variety to which he/she has been exposed to. As a consequence, the polysystemic representation of the native language will have then to interact with the formation and recognition of categories in the second language, which will also be constructed as a polysystemic representation of the mapping of phonetic details of the L2 with the experienced contexts of use.

The learner, in this case, must store a high quantity of information and has to revise his/her interlanguage system (as it will be described in section 1.5.) several times; every time that a new context is experienced, and that a new phonetic detail is noticed in relation to it.

Nonetheless, this perspective is undoubtedly addressing a core point of the debate around the processing of speech categories, and to some extent, it fits the debate on the nature of intonational categories.

1.3.8. Summary

In this section, I have addressed the process of categorization and the resulting outcome - the category. The function of a category is the possibility to group under one label objects sharing at least a common property (Harnad, 2005). The grouping process reduces the quantity of information that has to be stored in memory. After the categorization process, information, consisting of a plurality of objects, can be stored and recalled through the one single label after the categorization process (Robert, 2005). Once several categories are established, the information can be organized establishing relations among the available categories (Relation2-sameness), and this process allows to think causally and is related to the presence of language (Holoyoak & Thagard, 1995; Robert, 2005). The presence of categories and of the categorization process itself is relevant because the structuring of information allows connecting a high number of objects through inferential mechanisms. The ability to make inferences about a whole set of members of a category allows making predictions about the properties of an incoming piece of information, given the assumption that it is a member of it, too. On the other hand, the incoming piece of information is processed according to the relevant properties of the available categories, because it has to be stored according to that structure. Thus, the presence of a structurally organized set of categories creates expectancies on the properties of an incoming stimulus and sets an attentional template guiding the search for relevant features of the stimulus. The presence of structurally organized categories generates, therefore, a top-down activation affecting perception. On the other hand, at the start of the whole system, when categories have to be formed, the categorization process analyzes a high number of inputs to establish if there is a feature invariantly shared by many inputs, so the system starts analyzing bottom-up several incoming stimuli.

Research on speech perception and acquisition focusing on the segmental level has proposed several models seeking to account for the role of categories in the sound domain. In this section, I have addressed the CP paradigm (Liberman et al., 1957) that predicts the loss of discrimination between objects already grouped within one category. Secondly, I have presented the PME (Kuhl, 1994), that predicts the presence of prototypes attracting all categories members as perceptual magnets. The presence of magnets, contrary to what the CP paradigm assumes, leaves the possibility to discriminate within category-members, also when the category is already established.

Firstly, the debate about the presence of categories and language is guided by the presence of results indicating that not all sounds in language are equally categorically perceived (e.g., vowels are less categorically perceived than consonants). Secondly, moving on to the field of intonation, more than one factor affects perception. In many studies, the presence of CP has been sought during the perception of pitch rising or falling and aligned in sentence-final position (BTs), and perception has been defined as quasi-categorical. Quasi-categorical perception means that a fairly high discrimination sensitivity, above chance, is present during the perception of several objects that continuously vary ranging between two possible categories. The discrimination rate remains high not only in the presence of a category-boundary but also for a broader range of perceptual object. Moreover, that the category boundaries dividing the set of objects are less clear-cut. It could be said, that the prototypes of the two opposed categories, in terms of PME, are weaker in attracting the surrounding perceptual objects. Certainly, weaker than in the classic cases of consonants perception.

Nonetheless, as introduced in the previous sections, within a language system, the recognition of a rising movement labelled as L+H* (in AM), establishes a phonological representation possibly associated to the pragmatic act of asking a question, and requiring from the interlocutor an answer narrowly or broadly focused, semantically congruent and predictable from the content of the question. It is relevant therefore to understand how and whether the sound representation is recognized as a discrete category and how it maps with segmental, semantic, and discourse information.

The first relevant property is that the perception of pitch movements is affected by the rate of modulation. Faster transitions are recoded as relative pitch levels, while slower transitions are perceived as movements (House, 1990). The second relevant factor is the role of segments since pitch movements align with them. Vowels are the type of segments within the syllable that can be prolonged over time while maintaining stable spectral information. Thus, pitch movements aligned to vowels can be modulated over longer stretches, increasing the chances to be perceived as movements. On the contrary, the faster transitions between consonants, and consonants and vowels within the syllables favor the perception of pitch levels.

The Kiel Intonation Model (KIM) (Kohler, 1991, 1997) is centered on the role of tonal alignment. It has been adopted to test the role of alignment in the presence of peaks and valleys in the pitch contour, showing different results. It identifies 3 peaks alignment-positions with a phonological status: early, medial, and late, and their positions interact with the shape and rate of the modulation of the pitch contour. The perception of the alignment positions of valleys, on the other hand, has been investigated in interaction with the recognition of a pragmatic function (Niebuhr & Kohler, 2004). In this case, a clear-cut CP has not been found, and the interpretation of these results called for a perceptual-cognitive explanation. Valleys are characterized by the absence of a standing-out salient cue (contrary to peaks). The task to align a low pitch contour to a vowel and to identify it by shifting its position within the syllable, recognizing early vs. medial peaks, is harder when compared to the recognition to the peaks' position, because of the lack of anchor-points with the segmental information. Therefore, there is a shifting pitch modulation perceived not in a point in time but within a period of time, lasting for the duration of the vowel.

In absence of segments, but in presence of formant – information as the case of the adoption of the PRAAT "humming" condition, the recognition of the peak alignment position shows to be granted by the interplay of frequency and intensity modulation.

This perspective, on one side, challenges the concept of CP as classically conceived, and the role of abstract phonological representations guiding it. It points to the fact that, the pragmatic functions associated to pitch can be associated with two recognizable categories at the phonological level, but then the phonetic features that can be modulated and perceived as continua affect the perception tasks facilitating (in case of fast transitions and the presence of salient cues) or increasing the difficulty of the CP task.

The role of the phonetic details is centrally adopted by the Firthian Prosodic Analysis (FPA) and from the deriving polysystemic configuration of speech processing, the Polysp system (Hawkins, 2003). Within this perspective not only intonational categories, but speech in broader terms, is processed assuming a central role of phonetic details, and allows to overcome some of the inconsistencies that emerge by strictly dividing phonetic details from abstract phonological representations. On the other hand, such a perspective requires a high computational power, and to the author's knowledge, it is not fully clear whether such a computational power has been fully demonstrated in terms of cognitive processes and neural network activations, especially in cases in which more than one language system is activated.

The debate on the nature of intonational categories is relevant within this work because, firstly, in Chapter 2, the role of the recognition of segmental information will be tested through a classical categorization task. The presence and recognizability of segments can be read as the presence of recognizable anchor points, facilitating the identification of a specific pitch modulation, as well as the recognition of the pragmatic interpretation of the contour. Thus, the humming condition will be adopted requiring the identification of the category statement vs. question, not varying the identification procedure.

Secondly, in Chapter 4, the mapping between prosodic categories and information packaging entails a condition presenting a high acoustical saliency, through the presence of a peak, and a condition of low acoustical saliency that has to be recognized as incongruent by the property of lacking modulation. In Chapter 4, the recognition of an incongruency between pitch and information will be tested in native and second language speakers.

The perceptuo-cognitive perspective proposed by Niebuhr & Kohler (2004), signaling that the lack of perceptual cues in a specific point in time, decreases the emerging of CP will interact with the fact that Deaccentuation (to a certain extent given the ongoing debate on plasticity (section 1.2.)) shows not to be a frequent and non-marked strategy in Italian (native speakers of the L2 learners in Chapter 4). Thus, the mapping between the sound domain and the discourse structure in German L1 could be potentially based on a less categorically perceived phonological category – Deaccentuation – that indeed is not equally and similarly represented in the Italian L1 system.

As it will be introduced in section 1.5., differences in alignment strategies across languages have been considered in the formulation of an acquisition theory, but the degree of CP due to alignments' properties interacting with the L1 and L2 categories available to the learners has not yet been fully addressed. I will support the view in Chapter 4, that the presence of acoustical salient cues in the signal vs. the absence of acoustical salient cues interacts with the modulation of the response in ERP components, in function of the pragmatic structure.

In this thesis, pitch perception is investigated with methodologies providing information about the processing stages over time.

In the next section, I will address some evidence coming from studies adopting neurophysiological techniques that debate on pitch perception with linguistic and non-linguistic functions in speakers of tonal and non-tonal language adding information about the involved neural networks and specific areas. The debate on the neural networks involved in the processing of pitch can provide further information on the role of language and on how language affects the categorization process of pitch movements.

1.4. Neuroscientific evidences and models of pitch perception

There are two hypotheses in the neuroscientific field specifically regarding the processing of pitch: (i) the functional hypothesis predicts that the interpretation of pitch guides the neural pattern of activation; (ii) the acoustic hypothesis predicts that regardless of its function F0 is processed according to its acoustic features (for an exhaustive review, see Wong, 2002). Specifically, the functional hypothesis makes a general distinction on hemispheric specialization predicting left dominant activation for linguistic interpreted pitch stimuli and a right hemisphere dominant specialization for affective prosody. The acoustic position predicts that the right hemisphere is generally dominant in F0-decoding despite its function. Lesion studies and neuroscientific studies have been found to support both the hypotheses in a mixture of results and an unclear picture. Within this work I do not focus on the difference between linguistic and affective prosody per se, therefore I won't strictly support either of the two positions. Nonetheless, I briefly report here the main differences and results of the two theoretical perspectives on brain network activation in speech perception, speech recognition and auditory-motor mappings, carried out by different neural pathways of activation (Hickok & Poeppel, 2007). I report, as well, results of neuroscientific studies highlighting the role of language experience – top-down activation – and the role of the functional interpretation of pitch.

One of the most influential neurocognitive models of speech perception is the Dual Stream Model (DSM) (Hickok & Poeppel, 2000, 2004, 2007; Poeppel & Hickok, 2004). Despite the fact the DSM does not explicitly include a discussion of perception and categorization of intonation, it is worthwhile to briefly discuss some aspects described in the DSM in relation to the processes linked to categorization of speech. The name DSM refers to the definition of two main speech processing neural pathways: the ventral and the dorsal stream. In the ventral stream, the processing from acoustic analysis to semantic access is bilaterally computed; in the dorsal stream, the auditory-motor mapping between acoustic features and vocal tract configuration is computed, for single segments and sequence of segments (e.g. coarticulation); the dorsal stream processing is defined as strongly left lateralized and responsible for speech development and short-term phonological memory. Within the DSM, the two streams share a specific brain region: the left Superior Temporal Gyrus, that is at the core of dissociation phenomena found in patients and it is at the origin of the need for further neural pathways specification.

Firstly, the DSM is formulated to answer specific questions, emerged from non-converging evidence comprehending: (i) deficits reported by patients with lesions; (ii) the results obtained with early neuroimaging studies in the 1980s, and (iii) the, at the time, proposed theoretical predictions. In particular, the model proposed for speech perception, from roughly 1870s to 1970s, was based on the intuition that speech perception was supported by the auditory cortex, highlighting the role of the left Superior Temporal Gyrus (STG). The growing non-converging evidence, instead, began to report: that lesions to the left STG were causing impairments in speech production, but they were not compromising speech comprehension; and that deficits in perception of speech sounds appeared to be unrelated to comprehension deficits, as we understand them nowadays under the name of Wernicke's aphasia.
Moreover, evidence began to highlight the role of frontal and inferior parietal areas in left hemisphere in the identification or discrimination of speech syllables, which in patients appeared to be dissociated from word recognition.
Then, early functional neuroimaging studies highlighted that passive listening to speech activates bilaterally the superior temporal regions, whether syllable discrimination's tasks or analogous activate predominantly the left STG.

Therefore, the first strong fact that the early neuroscientific evidence confirmed is that "neural organization of speech processing is task dependent" (Hickok & Poeppel, 2007; p.393), meaning that different neural networks are activated, (and dissociated in patients displaying specific impairments correlated with lesions affecting one of two pathways), depending on whether the task requires to operate at the sound units, phonemes and syllable structure, or at the word level and semantic integration or access. The authors define speech perception, opposed to speech recognition, the processing concerning sound units in speech.

The first innovation, in comparison to previous existing models, is that the processing from the perception of raw acoustic signal to word recognition is assumed to occur through multiple parallels routes, bilaterally organized. Assuming multiple parallels routs supports the concept of a redundant processing system that makes use of redundant spectral information in speech, combined with temporal cues, and can explain limited impairments of single processing routes that do not compromise the whole perception system.

The second relevant observation made by the two authors is that most of the studies seek to individuate areas and network related to speech processing vs. non-speech processing, assuming for example, the logic of comparison of word processing vs. pseudo-word processing. Pseudo-words are thought not to be represented in the mental lexicon but to share with word processing the activation of the networks responsible for phonemes and syllable processing. In the same way, the authors argue, speech and non-speech single sound-units could share processing network responsible for the decoding the shared acoustic features and diverging in the representation or not representation in the phonemic inventory. With this perspective in mind a clear-cut division between speech and non-speech processing, disconnected from the experimental task requests at sound units' level, could not be necessary and neither functional to extend our understanding.

The third relevant factor defined by the model and extremely relevant because of the implications it has, is the processing's time scale and the neural pathways related to different time scales. Segmental information processing, intended as phonemes' recognition and phonemes' sequence recognition, requires a time window of 20-50 ms. Suprasegmental information - intended in the DSM as: syllable boundaries, syllabic rate, (lexical) tonal information, prosodic cues (for interpretations), and stress cues processing (parenthesis in authors' version, p. 396) – requires in the model a time window between 150-300 ms to be processed, corresponding to the envelope of the spoken utterances. The DSM predicts a multi-time integration where the information coming from the two time-scales is concurrently processed.
Functional neuroimaging results support the view that the integration of longer time scales – suprasegmental features – is predominantly right lateralized and information on short transitions – segmental level – is bilaterally processed. A further proposed asymmetrical differentiation pertains specifically the categorization process, as it is proposed that "the left hemisphere might be predisposed to processing or representing acoustic information more categorically than the right hemisphere" (Hickok & Poeppel, 2007, p.397)

In relation to this thesis, the DSM model highlights the role of the experimental task that can activate different neural routes depending on the level of analysis that it imposes – segmental, syllabic, lexical, sentence-level domain. Moreover, the experimental task in studies involving patients showed that a lesion within one route, does not necessary compromise the processing of information at a different level (sound discrimination vs. lexical access in aphasia patients).

The second fact is that speech and non-speech sounds can share hypothetically patterns of neural activations on the basis of common acoustic features, but as pointed out in the previous paragraphs the classification of specific sound pattern as speech units – phonemes, lexical tonal events, prosodic tonal events such as Pitch Accents – is language and experience dependent. Experience dependency can be defined in long-term stable mental representation of linguistic units, and relations among units, but also can refer to context dependencies as formed, or already existing but strongly activated, within and by the experimental setting. Degrees of categorization processes, or in classical terms differences between categorical or quasi-categorical effects in sounds categorization tasks could reflect then differences of activation patterns in term of neural routes since categorical linguistic processing is hypothesized to be more left lateralized.

Nonetheless, the DSM does not address, in a precise matter, suprasegmental aspects of speech, since it considers, within a non-further specified group, different prosodic aspects, syllable boundaries, syllabic rate, (lexical) tonal information, prosodic cues (for interpretations), and stress cues processing (parenthesis in authors' version, p. 396). In linguistic theory, segmentation and phrasing processes, rate of speech, rhythmic structure, and modulation of the fundamental frequency at lexical and sentence-level domain are considered separate phenomena, all converging under the term prosody. Thus, the DSM is under-specified compared to, for example, the theoretical formalization of the Autosegmental Metrical (AM) theory of intonation.

Moreover, from more recent neurophysiological evidences – e.g. ERPs studies – it is known that prosodic processing of phrase boundaries, for example, can require much longer time-windows than the ones hypothesized by the DSM, reaching the limit of 2000 ms (Steinhauer & Friederici, 2001) and that average processing of Pitch Accents at Discourse Level requires at least a time window spanning up to 800 ms (Schumacher & Baumann, 2010). It is conceivable that the hemispherical asymmetries described by Hickok & Poeppel (2007) pertaining the categorization process and its gradience, and the time-scale of inputs can be further specified in the future.

In the following paragraphs, some specific neuroimaging studies specifically on pitch are reviewed. They highlight the role of contextual information, experimental task, and native language in neural processing of F0. Unfortunately, as Wong (2002) and Wong, Parsons, Martinez, & Diehl (2004) reported, many lesion and neuroimaging studies do not report a crossed experimental manipulation. Many times, only one population of interest, e.g. Right Hemisphere damaged patients, is investigated, perceiving or producing one single experimental manipulation - e.g. the processing of affective prosody. Therefore, I choose to report only studies in which the experimental setting considers the full paradigm of experimental manipulations that , in my opinion, can have a higher explanation power of the correlation between the experimental manipulation and the neural pathway.

Some studies highlighting the role of language experience in processing of pitch patterns made use of lexical tones comparing perception of native speakers of tonal languages vs. speakers of non-tonal languages.
Wong et al. (2004) investigated with Positron Emission Tomography (PET) the perception of three out four tonal movements available in Mandarin Chinese, aligned on Mandarin Chinese syllables, creating in this way meaningful words in Chinese, on one side, and aligned to English words, on the other one; they presented the stimuli in pairs, matched by language in the segmental material, and asked participants whether the tone pattern within the presented pair was the same or different. Participants belonged to two groups:

native speakers of American English with no exposure to Mandarin Chinese, and native speakers of Mandarin Chinese also speaking English.
Behavioral results display higher accuracy in both stimulus types by native speaker of Chinese, and slower RT of both languages' groups participants performing the tone task on Chinese words than the pitch task on English words. In addition, Mandarin speakers showed to have a significant increase of regional Cerebellar Blood Flow (rCBF) in the left anterior insula and left basal ganglia when performing the task in Mandarin Chinese. Additionally, Wong et al. (2004) showed a bilateral activation in midline regions of the cerebellum in the posterior hemisphere, and right somatosensory motor cortex and premotor cortex. English speakers with the same stimuli displayed similar results but no cerebellar activity. Instead, English speaker intensively activated the right anterior insula, homologous to the activation of the left anterior insula found in Chinese speakers. In the task with English words aligned to equivalent pitch movements both Mandarin and English speakers showed mainly right hemisphere activations. Mandarin native speakers showed a strong activation of the right insula and right cerebellum. English speakers perceiving English words with pitch modulation highlighted a strong activation of the right insula. English speakers only in both tasks exhibited a bilateral activation of the parietal cortex not-directly connected with speech processing possibly interpreted as the searching of alternative strategies outside speech domain in order to compute the task, such as word visualization.
The Authors concluded that the results exhibited by Chinese native speakers support a functional processing of pitch, since when the exact same pitch movements belonged to the native language organization, the processing appeared to be strongly left-lateralized and in L2 the same discrimination task on pitch movements led to the activation of the homologous insula area, but in the right hemisphere. The activation of the left anterior insula seemed to be correlated with a lexical function of tonal movement and of the right anterior insula with a sentence-level or non-lexical function of pitch. The Authors did not exhaustively explain the role of the cerebellum, but they suggested, it is connected with auditory sensory processing and working memory access.

Xu et al. (2006) investigated with fMRI the perception of Mandarin Chinese tones superimposed on Chinese syllables (existing Chinese word - CC) and the perception of Thai tones superimposed of Chinese syllables (non-existing tonal chimera - CT), in native speakers of Chinese and native speakers of Thai. The behavioral results exhibited higher accuracy for CC contrasts by native speakers of Chinese, and higher accuracy in CT contrasts by native speakers of Thai. The perceived difficulty in computing the task was rated: Chinese speakers found significantly easier the task with CC contrasts and rated the task both with CC and CT comparisons to be equally difficult. Between groups comparisons indicated higher accuracy in CT contrasts in the native language speakers of Thai group, and CC contrasts were perceived significantly easier to be discriminated by the native language speakers of Chinese group. When comparing the activation found for CC perception over CT perception in native speakers of Chinese, and of CT perception over CC perception in Thai speakers, the Authors found overlapping brain activation in left Planum Temporale (PT). Examining each task separately (CC, CT), no significant laterality effect was found in either language group. Xu and colleagues interpreted the left PT activation in function of language experience, since the region correlated with the higher degree of experience with the respective language tonal modulation. They proposed that the left PT has "access" to discrete phoneme categories (quote present in authors' version, Xu et al. 2006 p.180). Interestingly they used the term phoneme when the comparison pertains only tonal movements coded differently in the two tonal languages. Left PT then showed to have the function of hub, matching top-down processing and bottom-up processing on incoming stimuli already at relatively early stage of acoustic-phonetic processing and that neural plasticity showed to be guided by language experience. In particular they concluded that “left PT appears to be sensitive to learning experience that involves the development of abstract cortical representations irrespective of cognitive domain or sensory modality” (p. 180). Regarding bilateral activation of PT, the examination of each condition suggested that pitch processing per se, independent from language experience, is responsible for the activation of the auditory cortex in the right hemisphere, as supported by acoustic hypothesis (Zatorre, Belin, & Penhune, 2002), whether the accessing to higher-order categorical representations shifts the processing to the left PT in language-sensitive tasks.

Neural pathways differences have been also found comparing sentence-level linguistic prosody to emotional prosody aligned to the same segmental information (Wildgruber, Ackermann, Kreifelts, & Ethofer, 2006; Wildgruber et al., 2004).
Wildgruber et al. (2004) aligned to neutral sentences in German (“Der Schal ist in der Truhe”, ‘The scarf is in the chest’) an F0 modulation that, in a linguistic condition, shifted from conveying a Narrow Informative Focus associated to the referent “der Schal” to a Narrow Informative Focus centered to the final word “Truhe”. The shift in focus position and F0 modulation – realized enhancing the peak on the word “Truhe” – changed in the experiment the interpretation of the sentence from being a congruent answer to the question “what is in the chest?” to a congruent answer to the question “where is the scarf?”.
In an emotional condition, pitch range was gradually raised homogeneously over the whole utterance as representing a raising of excitement of the speaker. In the linguistic condition, pairs of stimuli were presented, and the task consisted to respond to the question “Which of the two sentences is better suited to respond to the question: Where is the scarf?”. In the emotional condition, analogously, participants had to decide “Which of the two sentences sounds more excited?” (Wildgruber et al., 2004; p. 1384).
The results highlighted that the extraction of acoustic pattern required to make pitch judgments independent from condition correlated with bilateral, but partially rightward lateralized, activation within the dorsolateral frontal cortex, frontal operculum, anterior insula and superior temporal cortex.
On the other hand, the linguistic condition predominantly activated the left Inferior Frontal Gyrus (IFG) and the emotional condition strongly correlated with bilateral orbito-frontal regions. The activation of the left IFG was proposed to reflect the focus shift processing involving syntactic and pragmatic function of pitch movement in the linguistic condition.

Further evidences of differential processing of auditive stimuli involving full linguistic activation and not-full linguistic activation came from Meyer, Steinhauer, Alter, Friederici, & von Cramon (2004). In a first experiment, regular sentences and pseudo-speech sentence (in which the syntactic structure and the functional words were preserved) were auditorily presented to participants. Both regular speech sentences and pseudo-speech sentences correlated with stronger left lateralization of Superior Temporal Regions (STR) and they both are associated to bilateral activation of thalamus. Pseudo-speech sentence in addition displayed to correlate with bilateral activation of the deep frontal operculum, close to the anterior insula. In a second experiment, Meyer and colleagues compared the processing of normal speech, pseudo-speech and degraded speech, where segmental information was filtered out and only F0 modulation remained available to perceiver. In both experiments, the task for participants was to state whether the perceived sentence was produced in an active or passive verbal voice, information impossible to grasp in the degraded speech condition. The results of the second experiment showed a bilateral activation of the IF lobes and ST lobes for both normal and pseudo-speech. All 3 conditions shared the association to primary auditory cortex and supratemporal planum. In addition, normal speech displayed a similar pattern of activation as in the first experiment; pseudo-speech, compared to normal speech, highlighted a bilateral increased activity in the STR and FOR. The degraded speech condition compared to the other two conditions showed an attenuation of the STR activation and an enhanced activation of the right fronto-lateral areas (IPS) and a bilateral activation of the front-opercular cortex. The Authors suggested that the stronger activation of STR in pseudo-speech conditions in both experiments reflected the linguistic processing up to the point where lexical entries were found missing, but the syntactic and morphosyntactic aspects were correctly processed. The frontal activation, stronger in the right hemisphere found for the degraded speech was suggested to reflect the increase of effort related to the segmentation task; in degraded speech, in fact ,the segmental information supporting word segmentation, was missing.

In summary, neuroscientific evidence reported three major relevant facts: there is consistent evidence supporting the acoustic hypothesis of a significant role of the Right Hemisphere in the processing of slower frequencies modulations, independently from their linguistic and emotional functions, as described by the DSM.
This can suggest, for example, in Categorical Perception (CP) paradigms adopting vowels and intonational - stimuli, that the substantial difference encountered with respect to the CP of consonants (emerging in a gradient modulation of the identification function and in the presence of a plateau in the discrimination task) can be due to partially different pathways of neural processing, predominantly involving the Left Hemisphere in case of fast frequency transitions (consonants) and the Right Hemisphere in cases of slower transitions (pitch and vowels).

In second instance, pertaining pitch processing, language experience appears to play a determinant role on the gradience of discreteness that can be associated to single F0 movements. Language experience has been shown to be able influence the neural plasticity modifying early stage of acoustic processing.
Therefore, there is not probably a single description of the discreteness associated with pitch movements and a single possible prediction of outcome in CP paradigms.

Thirdly, at sentence-level domain, pitch movements associated to focus structures showed to be able to shift the neural processing involving left hemisphere linguistic networks, whether emotional prosody did not. Observing only the signal properties and ignoring contextual information, both discourse level manipulation and emotional manipulation of perceived excitement involved the partial or general rising of F0 peaks; a change in processing networks cannot be then only ascribed to the acoustic properties of the signal but instead it becomes clear that the contextual information, e.g. presence of segmental information, as well as the experimental task, guide a process of pitch interpretation as linguistic or paralinguistic cue at sentence level.
Moreover, pitch can be integrated with segmental information in normal speech sentences and pseudo-speech sentences but differences in STR activation have been found in function of full or absent semantic integration or access. In final instance, a higher cost consistently corresponding to a bilateral and right hemispheric frontal regions activation highlights that pitch processing is probably computed with the aid of the syllabic structure and word-segmentation cues, that are missing in degraded speech.

1.5. The Second Language Acquisition of speech categories: The challenge of modelling the acquisition of intonation.

In this section, I address the fundamental concepts adopted in Second Language Acquisition (SLA) as well as the main results explicitly pertaining to the acquisition of L2 pitch categories. The section will end presenting the current ongoing debate and the recent proposal of the L2 intonation learning theory (Mennen, 2015).

I would like to highlight here, how the debate on speech processing in L2 started adopting the fundamental concepts: the role of the age of acquisition, the construction of an interlanguage system, the effect of the presence of the native language, and the structural differences between the native language and the L2. Then, models that assumed the presence of either phonetic or phonological categories (e.g., the Speech Learning Model (SLM) and the Perception Assimilation Model/L2 (PAML2) (Flege, 1995; Best & Tyler, 2007) entered the debate, focusing on the acquisition of segmental information. The debate, then, shifted to the acquisition of intonational categories, adopting the AM framework, formalizing the interplay of 4 dimensions (systemic, realizational, semantic and frequency). The issue, therefore, of understanding the nature of categories within the sound domain and their interplay with the other dimension, as well as the debate on the phonological or phonetic nature of alignment remains. The picture becomes even more complicated since the investigation here requires to consider the age of acquisition, the structure, and categories of two language systems, and the organization of the discourse in two language systems.

In the following paragraphs, firstly, I address the role of the age of acquisition, the construction, and evolution of the interlanguage, and how different language properties, differently appearing in native and non-native languages, affects the acquisitional process.

1.5.1. **Fundamental concepts in Second Language Acquisition**

1.5.1.1. Critical Period

One of the most debated concepts in the field of Second Language Acquisition (SLA) is the Critical Period Hypothesis (CPH) (Bialystok & Hakuta, 1994; Birdsong, 1999, 2006; Curtiss, 1977; Lenneberg, 1967).
The CPH of Lenneberg postulates that there is a specific period, corresponding to infancy, in which language can be automatically learned, but that period of time terminates approximately with puberty. Age, thus, affects the automatic acquisition of language and once the learner grows older than the critical period, the acquisition has to occur through conscious effort and formal instruction The CPH affects directly first language acquisition and secondarily has implication for the field of SLA. The cases of first language acquisition during or after puberty are rather few and connotated by exceptional circumstances and conditions. Probably the most famous case is the one of Genie, found at the age of 13 after being isolated and segregated for the whole life, with no social interaction, and therefore without having acquired any native language. Genie demonstrated the ability to acquire verbal and non-verbal linguistic skills, she could, for example, acquire new words and improve her lexicon, but she never properly native-like acquired some morpho-syntactic aspects of English and never learned how to correctly produce interrogative and wh-questions as well as complex sentence structures involving subordination. Her intonation has been reported as "odd" by native language speakers. Yet, the case of Genie strictly did not support the strong original version of the CPH, since she somewhat could begin to acquire her native language later in life. The CPH, in this case, seemed more to reflect the quality, or a degree, of acquisition of the native language, rather than

a 1/0 condition. Another case of late acquisition proved that L1 acquisition could successfully occur even at the age 18 (Buddenhagen, 1971).
The CPH has also been tested in SLA showing that the area where it mostly applied, regarded the acquisition of the phonological system and the linked phonetic realization, causing an L2 production characterized by Foreign Accent (FA). (Hyltenstam & Abrahamsson, 2000, p.152) summarize in the following 5 points the implications overall descending from the CPH:

1. Younger language learners are `better' in second language learning than older learners.
2. Younger learners outperform older learners with respect to eventual outcome.
3. Younger learners acquire second languages automatically from mere exposure, while older learners have to make conscious and labored efforts.
4. Younger learners can reach native-like levels of proficiency, but older learners cannot.
5. The turning-point age for differences between children and adults in 1 to 5 is puberty.

Since the 70s several studies aiming to support or disprove the CPH have been performed, generating an unclear picture. A sequence of studies supported the claim that younger learners outperform older ones. The disadvantage of late learners has been explicitly reported for pronunciation (Munro, Flege, & MacKay, 1996; Oyama, 1976; Tahta, Wood, & Loewenthal, 1981) and grammar learning (e.g. Coppieters, 1987; Harley, 1986; Harley & Hart, 1997; Johnson, 1992; Johnson & Newport, 1991; Patkowski, 1980; Schachter, 1990; Sorace, 1993).
One of the most cited study, Johnson & Newport (1989) tested the acquisition of syntax and morphology in L2 learners of English, L1 speakers of Chinese or Korean, arriving in the United States. They divided the participants into 2 groups early (age 3-15 years of age) and late arrivals (17-39 years of age). Their main finding is that the age of arrival linearly predicts the accuracy in the scoring of grammatical tests, and also that the linear trend is strongly reliable in the group of early-arrivals, but on the contrary, it displays greater variability, thus the presence of individual differences, in the group of late-arrivals. In second instance, they statistically observed whether the beginning of formal instruction could explain further variance in the data when compared to the only criterion related to the age of arrival. Nonetheless, they do not find any significant effect, supporting, therefore, a stronger role of the immersion factor, when compared to the presence of formal instruction.
On the contrary, a similarly conspicuous amount of studies found results going in the rather opposite direction and therefore does not support the CPH as originally proposed (see: Bialystok, 1997; Bialystok & Hakuta, 1994, 1999; Birdsong, 1992, 1999; Bongaerts, Planken, & Schils, 1995; Bongaerts, Van Summeren, Planken, & Schils, 1997; Flynn & Manuel, 1991; Krashen, Long, & Scarcella, 1979; Long, 1990).
One of the earliest works, showing evidence disproving the CPH is Snow & Hoefnagel-Höhle (1978). In this study, the authors tested the acquisition of several aspects, in English L1 speakers acquiring Dutch, without formal instruction, for one year. The participants were longitudinally tested during the first year of permanence in the Netherlands. They tested 51 participants divided into 5 groups of age (3-5, 6-7, 8-10, 12-15, adults). The participants were measured in 3 points in time and compared with a control group of native speakers and a group of advanced L2 learners of Dutch. The participants were tested in all language components, and extensively in the acquisition of pronunciation and auditory discrimination. This study is non-supporting the CPH since the main results is an acquisitional advantage of the age group between 12-15 years of age. In particular, younger children between 3-5 were outperformed in all tests, and this result has been taken as main evidence disproving the CPH. This seminal study is also interesting since it is one of

first already reporting a significant presence of individual variability especially in the pronunciation and auditory discrimination tasks.
All together the empirical evidence led to the conclusion that the L2 acquisition in function of age is not a monolithic phenomenon, but each area of investigation is differently affected by the age factor. From the first group of studies to the second one, the description and the factorization of sample properties and linguistic aspects has been refined. This led to the proposition of the existence of more critical periods, depending on the area (Diller & Walsh, 1981; Eubank & Gregg, 1999; Long, 1990; Schachter, 1996; Scovel, 2000; Seliger, 1978). A clear advantage of early acquisition has explicitly been found for phonological aspects and the strong version of the CPH, which presupposes that after puberty acquisition stops, has been replaced by a softer version proposing that the possibilities of a speaker to ultimately sound native-like in a second language decline along with age, but the ability to acquire does not abruptly vanish away at a specific age (Bialystok & Hakuta, 1994, 1999; Birdsong, 1999; James Emil Flege, Yeni-Komshian, & Liu, 1999).
A further step in explanation came with Bley-Vroman's (1988) Bley-Vroman’s (1988) proposal of the Fundamental Difference Hypothesis (FDH). The FDH aims to explain the differences in acquisition between early and late learners hypothesizing that the main difference between the two groups is the possibility to rely on the innate mechanism responsible for language acquisition. Late learners can no longer rely on it and therefore, have to adopt alternative mechanisms. These alternative strategies are verbal-analytic problem-solving skills combined with a fully developed cognitive system. Together these aspects can explain why in some tasks, adults can outperform children in the short time. The cases of older adults able to successfully acquire an L2 system with native-like performances are indeed present (e.g. Abrahamsson & Hyltenstam, 2008; Birdsong, 2007; Bongaerts, 1999; Colantoni & Steele, 2006; Reiterer et al., 2011; van Boxtel, Bongaerts, & Coppen, 2005) and if the FDH accounts for a significant explanation of the differences then exceptional abilities in late-later have to be explained assuming that a high variability across individual is present, so that some individuals can overcome biological constraints.

1.5.1.2. Interlanguage

The second fundamental concept in the field of SLA is the notion of Interlanguage (Selinker, 1972).
Interlanguage refers to the systematic knowledge of a second language by an L2 learner, whose mental representation is independent of both the native language's representations and the L2's target ones: It refers to (Ellis, 2008, p. 968):

1)the series of interlocking systems which characterize acquisition;
2)the system that is observed at a single stage of development;
3)particular L1/L2 combinations.

The notion of interlanguage had a revolutionary impact in the analysis of L2 acquisition since it is strongly related to the idea that the acquisition of an L2 has an evolutionary process, and along the process, the interlanguage's representations vary, improving as the systematization of the new system occurs.
The analysis of the interlanguage comprehends, therefore, the ultimate analysis of a performance in a determined point in time along the acquisition, but also considers the cognitive aspects of the learner that proceeds in the acquisition both adopting common mechanisms as well as his/her systematization strategies. The interlanguage representation, thus, is independent of both the L1 and L2, but its construction

then is strongly dependent on the resources available to the learner, that he can use to start processing and memorizing a new linguistic system.

The evolution of the interlanguage towards the target L2 representations is a process of continuous restructuring of the available representations. Nonetheless, it can ultimately stop and never reach a native-like state. The acquisition can fossilize at certain stages. Fossilization is defined as (Selinker & Lamendella, 1978, p. 187):

> "…a permanent cessation in learning before the learner has attained target language norms at all levels of linguistic structure and in all discourse domains despite the learner's positive ability, opportunity, and motivation to learn and acculturate into target society."

As well as for the notion of CPH, the concept of fossilization has also been strongly debated, and ultimately reduced to a softer version of stabilization (Long, 2008). The stabilization concept is ground on the same assumptions of fossilization but assumes that in cases where the acquisition stops reaching a certain stability of the mental representations, it can ultimately again restart later, if the learner is motivated in achieving native-like speech performances.

The ultimate advantage of early learners, thus, can be inversely related to the fossilization/stabilization process since what is has been repeatedly found is the advantage in acquisition they have is the long run. Krashen et al. (1979, p.161):

> "-adults proceed through the early stages of syntactic and morphological development faster than children (where time and exposure are held constant);
> -older children acquire faster than younger children (again, in the early stages of syntactic and morphological development where time and exposure are held constant);
> -acquirers who begin natural exposure to a second language during childhood achieve higher second-language proficiency than those beginning as adults."

The late acquisition appears to take advantage of the developed cognitive system, well-formed L1 representations and to make use of this in a sort of comparative analysis with L2 signal while generalization processes occur. In the long run, nonetheless, late learners appear to more frequently slow down, or stop the acquisition process, once the available representations are sufficient for the communication.

Early learners, going through constant training, appear to maintain the ability to improve the interlanguage's representations, and be less affected by the fossilization/stabilization process.

1.5.1.3. Transfer

The third fundamental concept in SLA is the notion of transfer and interference. The notion of transfer has its origin in the behaviorist tradition of acquisitional theories. Nonetheless, it has remained within the SLA lexicon, and it refers to Ellis (2008, p.351):

> "Language transfer refers to any instance of learner data where a statistically significant correlation is shown to exist between some feature of the target language and any other language that has been previously acquired."

The most significant problem in the notion of transfer is that in the behaviorist terms the ultimate L2 performance of participants was taken as evidence reflecting the cognitive process. The strict correlation

between previously acquired languages and a determined L2 performance is not sustainable, given the independent nature of the interlanguage stage. On-line processing techniques such as the use of Event-Related Potentials (ERP) can allow overcoming some of the problematic aspects that are traditionally reported in transfer analysis since they allow to directly compare whether the processing stages in L1 and L2 learners different and not to ultimately rely only on performances/production data.
Traditionally transfer has been analyzed either comparing the production of i) one single feature in L1 and interlanguage and native speakers of the L2's systems; ii) speakers of several L1 acquiring the same L2; iii) bilinguals vs. L2 learners; iv) crossed designs where speakers of two L1s are reciprocally tested in the other language as L2. The traditional way with which transfer has been analyzed is the Error Analysis (Richards, 1974). The analysis of errors, nonetheless, has been disfavored since the introduction of the concept of interlanguage, that rehabilitates the concept of error as the particular category or rule representation that the learners have in that specific moment of the target language. Adopting the notion of interlanguage, errors can be organized in errors reflecting: the interlanguage developing process, the interference of L1 fixed mental representations and structures, and not reflecting any of the two factors but representing unique strategies of the learner. The transfer is not only a process causing the proliferation of errors, but it can also facilitate the acquisition: positive transfer. According to Ringbom (2007), learners do not ultimate seek differences between the languages but shared properties between the two systems. Starting levels of congruency between the languages facilitate the early stages of learning allowing the use of the assumptions given by the native languages, working similarly in typologically similar languages. This rather intuitive perspective, nonetheless, seems to function in the opposite direction when the acquisition pertains the phonological level. The issues will be addressed in the following section, but briefly, what the theories about the L2 acquisition of phonology strongly state is that a high similarity between the systems increases the difficulties in noticing phonetic details, and perception remains anchored to the perceptual categories of the L1, blocking the process of acquisition.

1.5.1.4. Markedness

The concept of markedness is defined for the first time within the school of Prague within the field of phonology (Jakobson, 1941; Troubetzkoy, 1939). Troubetzkoy formalized in those years the notion of distinctive feature and the following binary system that allows identifying one single phoneme at the time. Each phoneme in this way can be described by a combination of presence or absence of specific features (e.g. [±dental], [±voiced]). From this theoretical basis, two concepts emerged: simplicity and naturalness. Simplicity refers to the number of features necessary to identify a single phoneme. Fewer features signal a higher phonological simplicity. Naturalness refers to the property of some phonemes to be present in the phonological systems of all the attested spoken languages, e.g. /a/. The concept of naturalness is based on the notion of natural class. Two or more segments become a natural class if the number of features necessary to describe the class is smaller than the number necessary to identify a member of the class (Hyman, 1975). This leads to the implication that in order to describe a group of phonemes such as /b,d,g/ it is necessary to use 3 distinctive features: [+voiced, -continuous, -nasal]. On the contrary, the description of the following list of phonemes /b,d,g,v,z,m,n,l,r,w,j,a,e,i,o,u/, is possible by using only a feature [+voiced]. Following the description above, then, the second group of phonemes could be judged as more natural than the first one. Nonetheless, it is rather hard to apply any phonological rule to a group of phonemes such as the second one. It follows therefore that the simpler a class is, the less natural; nonetheless, it follows also that the more natural a phoneme is, also, the more frequent (e.g.,/a/ is more frequent than /i/ and /u/

across languages, thus /a/ is more natural). This generates what it is called an implication-universal, that states that if a language displays less natural phonemes such as /i/ and /u/, it must, therefore, display the more natural ones, such as /a/; on the contrary, if a language displays a relative more natural phoneme, it does not imply that a less natural one is also available in the inventory. To avoid confusions of terms, the property of /a/ with respect to /i/ and /u/ is expressed in terms of *markedness*. Trubetzckoy applies the notion of markedness expressively for the cases of neutralization (E.g., in German, the group of phonemes /b,d,g,v,z/ in word-final position is always produced as /p,t,k,f,s/. The distinctive feature [±voiced] is then neutralized, and therefore the two classes of phonemes are not distinguishable anymore by this feature).
Neutralized class of phonemes, thus, are less marked, compared to non-neutralized one.
Markedness is, then, adopted in the studies of typological differences between languages (Greenberg, 1966b), creating the concept of *typological markedness*, defined as (Gundel, Houlihan, & Sanders, 1986; p.108):

> "A structure X is typologically marked relative to another structure Y, (and Y is typologically unmarked relative to X) if every language that has X also has Y, but every language that has Y does not necessarily have X."

Markedness applies to SLA for what pertains the order of acquisition of specific features in native and second language acquisition. Non-marked elements are acquired before marked elements in L1, and this has led, by analogy, to propose typological markedness as an indicator of the difficulty of acquisition.
According to (Eckman, 1977, p. 321) the Markedness Differential Hypothesis (MDH) assumes that:

> "Those areas of the target language which differ from the native language and are more marked than the native language will be difficult;
>
> The relative degree of difficulty of the areas of difference of target language which are more marked than the native language will correspond to the relative degree of markedness;
>
> Those areas of the target language which are different from the native language but are not more marked than the native language will not be difficult."

The core of the concept of typological markedness is the third assumption because it generates a prediction in terms of successful acquisition, and it lays on one oriented vector the concept of difficulty of acquisition. The MDH has been tested over the years: Moulton (1962) tested the acquisition of the case mentioned as example before, thus the production of neutralized voiced plosive and fricatives in word-final position in German, supporting the MDH and showing that it is more difficult for German learners of English to correctly produce two distinct set of phonemes in word-final position in English, than it is for English learners of German to acquire one set of neutralized ones in German L2. Anderson (1987) tested the acquisition of consonant clusters in onset and coda of the syllable, formed by 2 phonemes. The hypothesis is that learners displaying consonant clusters of more than 2 phonemes in their L1 would be advantaged in the acquisition in L2, compared to learners not having a more marked feature in this respect in their L1. The results support the hypothesis highlighting that learners, Arabic L1, are facilitated in the acquisition compared to L2 learners of two different varieties of Chinese. An analogous result is also found by Carlisle (1991) for the acquisition of CC e CCC onset in English L2, by L1 speakers of Spanish where the most frequent syllable structures are CV and CVC. Supporting the hypothesis, L1 speakers of Spanish introduce an epenthetic vowel to facilitate the production in L2, confirming that for them the CC and CCC structures are more marked and difficult to acquire.

The MDH received some critics since it remains unclear what type of predictions are generated in the case both L1 and L2 display marked features concerning a similar phenomenon, and how it explains the cases in which speakers, coming from a marked feature in L1, remain resistant to acquire a marked feature in L2. An independent criterion of markedness fails to explain the possible relationship between the two languages. For these reasons Eckman (1991) revised the hypothesis, formulating it to include the concept of interlanguage, independent from L1 and L2, and assuming a relative perspective. The MDH became then a particular case, or a more general Structural Conformity Hypothesis (SCH) and the predictions generated with the concept of markedness implies that a more marked feature will always be more difficult to acquire than non-marked one and that it equally applies in interlanguage systems as well as native ones. More marked features are then assumed to be always more difficult to be correctly mastered within an interlanguage compared to the L1.

Recently in the field of SLA of intonation, the MDH/SCH approach has been used to test the acquisition of intonational patterns in L2 in Germanic languages that imply Deaccentuation in function of the pragmatic structure of discourse, a “feature” that appears to be more marked when compared to the intonational systems of Romance languages. Supporting the MDH it appears to be easier for native speakers of Germanic languages to acquire patterns of intonation that are guided by the metrical hierarchical structure and therefore displays the NPA in a predictable position, than for native speakers of Romance languages to consider both structural and pragmatic constraints to assign prominence in L2 (Avesani, Bocci, Vayra, Zappoli, & Chini, 2015; Rasier & Hiligsmann, 2007; Swerts, Krahmer, & Avesani, 2002). This point will be at the center of the EEG experiment involving L1 and L2 speakers of German and it will be extensively reviewed in the introduction of chapter 4.

1.5.2. The acquisition of the L2 phonological system

In this section I describe two accounted models of perception in L2: The Speech Learning Model (SLM)(James E. Flege, 1995; James E. Flege, Schirru, & MacKay, 2003; James Emil Flege et al., 1999) and the Perceptual Assimilation Model, in its latest version accounting for L2 perception (PAM/L2) (Best & Tyler, 2007). The two presented models provide an explanation and predictions on how categories present in the L2 are affected by the presence of the native language phonological system and therefore affect both the categorization and discrimination process. The two models do not come from the same theoretical perspective. The SLM emerges from research on SLA and the investigations pertaining to the presence of Foreign Accents in late learners. The primary interest is understanding how the fundamental concepts of L2 acquisition experimentally emerge in learners, and to provide an explanation of which variables precisely affect the acquisition. The PAM and PAM/L2 models derive from a specific position in the theory of speech perception. The aim of the theory is broader; it is interested in modeling primarily L1 acquisition, and into furnishing assumptions and predictions on the processing of speech. Nonetheless, Best and Tyler (2007) have reviewed both models and discussed similarities and incompatibilities and the PAM/L2, and it appears that they intend the PAM/L2 model has a complete description, but that both models do not substantially disagree on fundamentals assumptions. Together, therefore, they represent probably the most exhaustive formalization of how non-native categories are perceived.

The SLM is formulated in a period in which most of the research on L2 was performed adopting the Contrastive Analysis (CA). The CA presupposes that the structures and phonological inventory available in the native language of a speaker are compared to the structures he/she has to acquire. Adopting the notion of transfer and a technique called Analysis of Errors, the errors that the L2 speakers commit when attempting to speak in L2 are then attributable directly to the structures available in his L1. The CA predicts that L2 phonemes similar to the one available in the native language of the speaker will be easily reproduced; while on the other hand, phonemes that are different from the native one will be difficult to produce. The predictions of the CA hypothesis began to be doubted because of the evidence coming from experiments. Firstly, Flege (1987) found that adult L2 learners had an advantage in the acquisition of vowels that were not present in the native phonological inventory, compared to the acquisition of L2 vowels that "resembled" the ones available in L1. Secondly, also the hypothesis that only the abstract mental representation plays a role on acquisition has been disconfirmed by evidence highlighting that a critical factor was the production of phonetic features (Flege & Port, 1981; McAllister, Flege, & Piske, 2002). The SLM then formulated specific hypotheses on the L2 acquisition, that are centered on the role of the phonetic features of phonemes. It is based on 4 postulates (Flege, 1995):

- The mechanisms and processes used in learning the L1 sound system, including category formation, remain intact over the life span and can be applied to L2 learning.
- Language-specific aspects of speech sounds are specified in long-term memory representations called phonetic categories.
- Phonetic categories established in childhood for L1 sounds evolve over the life span to reflect the properties of all L1 or L2 phones identified as a realization of each category.
- Bilinguals strive to maintain contrast between L1 and L2 phonetic categories, which exist in a common phonological space.

And 7 hypotheses:

(H.1) Sounds in the L1 and L2 are related perceptually to one another at a position-sensitive allophonic level, rather than at a more abstract phonemic level.

(H.2) A new phonetic category can be established for an L2 sound that differs phonetically from the closest L1 sound if bilinguals discern at least some of the phonetic differences between the L1 and L2 sounds.

(H.3) The greater the perceived phonetic dissimilarity between an L2 sound and the closest L1 sound, the more likely it is that phonetic differences between the sounds will be discerned.

(H.4) The likelihood of phonetic differences between L1 and L2 sounds and between L2 sounds that are noncontrastive in the L1, being discerned decreases as Age of Learning increases.

(H.5) Category formation for an L2 sound may be blocked by the mechanism of equivalence classification. When this happens, a single phonetic category will be used to process perceptually linked L1 and L2 sounds (diaphones). Eventually, the diaphones will resemble one another in production.

(H.6) The phonetic category established for L2 sounds by a bilingual may differ from a monolingual's if: a) the bilingual's category is "deflected" away from an L1 category to maintain phonetic contrast between categories in a common L1-L2 phonological space; or b) the bilingual's representation is based on different features, or feature weights, than a monolingual's.

(H.7) The production of a sound eventually corresponds to the properties represented in its phonetic category representation.

The Perceptual Assimilation Model is framed within the direct realist approach to perception. The direct realist approach specifically assumes that both levels – phonological and phonetic one – are in direct relationship with articulatory gestures without the mediation of abstract representations. The realist approach in its foundation takes a specific theoretical position, which is set apart from the MTSP (nonetheless sharing some basic concepts as the role of the gestures and coarticulation in speech), but mainly, it sets a domain that is apart from the most classic position of generative grammar on one side, and apart from purely psychophysical approaches to perception that do not assume any role of the motor activation (for an extensive review, see Best, 1995). The specific aim to shift from models that investigate practically only production results; thus, the performance level, making at the same time claims on the phonological system of the learner, abstract competence level, to models accounting directly for the perception of the signal. In these terms, the perspective is almost the opposite to the one adopted by the Contrastive Analysis, since the interest, shared with the MTSP is to understand how phonetic features are mapped in phonological competence. The starting assumption is, also, that the phonetic implementation is linguo-specific, therefore native language dependent.

The developing of speech within this framework occurs through a process called "attunement," that involves the capacity to detect critical contrasts in the native language and discard irrelevant ones. It shares with the framework investigating statistical learning in infants that perceptual learning has to do with the ability to detect invariance and organize hierarchically. Lower-order invariants support the detection and organization of higher-order invariants.

The approach is centered on the idea that the detection of invariant corresponds to the detection and mapping of the articulatory gestures and patterns of coordination of the vocal tract. This specifically sets a restricted domain of what can constitute "speech," since the anatomical configuration of the vocal tract allows a defined set of possible configurations. Native language reduces the number of possibilities of configuration since in each language phonological inventories only specific contrasts are present. Perceptual attunement then involves the recognition of the relevant invariants and generates a prediction on what invariants have to be sought in speech perception.

Abstracting from the specific role of gestures in the model, the underlying mechanism of perception is in line with what it has been previously described, and again presupposes the development of selective attention mechanisms that "seek the linguo-specific invariant patterns" in the stimulus.

In particular, the model predicts that once the native language is acquired, attention is guided primarily by the phonetic domain of the L1 (low-order invariants), and the organization of critical L1 invariants becomes the phonological space.

> "Metaphorically speaking, the gestural constellations that serve phonological functions in the native language become easily recognized as well-worn paths in the spatiotemporal terrain of the universal phonetic domain" (Best, 1995; p. 187)

The application of this model to L2 acquisition then assumes that the phonological system of the native-language becomes the referring point for the detection of non-native invariants. The perception, in these terms analogously to Flege's hypotheses, is grounded upon the comparison between the L2 and L1 and the finding of similarities or discrepancies. Also, the predictions are modeled on the nature of gestures.
So, if the native language does not display any dental stop, but displays velar, bilabial and alveolar stops, and the L2 has dental stops in the phonological inventory, then the PAM predicts that L2 dental stops will be reconducted to the alveolar configuration since it is the configuration present in L1.
It formulates, therefore, a criterion of articulatory spatial proximity. L2 configurations closer to the native ones will be perceived adopting the L1 configuration, configurations classified as distant will be perceived as discrepant from the native ones, extremely unexpected configuration can be generally classified as "language" without a further specification.
The concept of categorization, therefore, here guides the process of assimilation of non-native phones. (Best, 1995).
A sound thus can be (p.194):

- assimilated to a native category;
- assimilated as uncategorizable speech sound, reflecting properties of more than one category in the L1 systems;
- not assimilated to speech, because too deviant from the constellations available in L1;

The model then predicts the following discrimination sensitivities to L2 sounds:

- Only one L2 phonological category is perceived as equivalent (perceptually assimilated) to a given L1 phonological category. Discrimination is expected to be excellent.
- Both L2 phonological categories are perceived as equivalent to the same L1 phonological category, but one is perceived as being more deviant than the other. Discrimination is expected to be moderately good.
- Both L2 phonological categories are perceived as equivalent to the same L1 phonological category but as equally good or poor instances of that category. Discrimination is expected to be poor.
- No L1-L2 phonological assimilation. Discrimination is expected to be either poor or good depending on the degree of dissimilarity between the patterns in L2 and L1.
- Categories are perceived as non-speech. Discrimination is expected to be good.

Best & Tyler (2007) provided, also, the following complementary comments to the postulates of the SLM.
Commenting the first postulate pertaining the possibilities to form new categories in the course of life, therefore outside the critical period, both models assume that during lifespans individuals keep refining their speech perception abilities. They both assume that environment, intending the languages to which we are exposed, influence the direction of perceptual learning and therefore the perceptual spaces of the categories the perceiver has. They specify, nonetheless, that categories that are formed in L2 and later in life are not assumed to be learned through the same mechanism of infants in L1. So, the PAM/L2 considers Age, and the presence of L1 a modifier of the perceptual strategies. In second instance, they strongly highlight the role that the articulatory gestures have in the PAM/L2 model, which are not present in the SLM, much more acoustically driven.
To the second postulate, that states that speech sounds are stored in long-term memory representations, the PAM/L2 opposes the concept of perception of articulatory invariant described above. Therefore, what

mainly changes is the conceptual notion of phonological and phonetic representation. In PAM/L2, a phonological category is information granting the discrimination of lexical differences. The phonetic category is the

> "invariant gestural relationship that is sub-lexical yet still systematic and potentially perceptible to attuned listeners. […] Such phonetic differences do not signal lexical distinctions but may instead provide perceptual information about the speakers' identity, or their region or language of origin." (Best & Tyler, 2007; p.21)

The third postulate, assuming that phonetic categories of the L1 evolve reflecting properties of any phone, of any language with which the learner came across, that has been classified as a realization of that category, is rejected in the idea that L2 sounds are perceived only as phones and not phonemes (the PAM/L2 model postulates the formation of a phonological level also in L2). Moreover, the PAM/L2 strongly accounts for the mechanism that selectively seeks articulatory patterns considered relevant invariant in L1, therefore predicts that irrelevant phonetic details are not processed. Nonetheless, it is again supported the fact that perceptual attunement can proceed along the lifespan.
The third postulate in SLM model, nonetheless, attempts to specifically account for influences of an acquired L2 on the structures of the L1. This is probably still addressed by the notion of assimilation in the PAM model, but perhaps it is not specifically addressed in terms of attrition.
In final instance, the 4th postulate, highlighting that there is a cognitive cost in keeping two linguistic systems in a common phonological space, is accepted in the main idea; nonetheless, predictions diverge since the PAM/L2 assumes a phonological representation of the L2. Therefore, in the PAM conceptualization, the phonetic information of the L2 is maintainable if found sufficiently distant from the configuration in L1.

In summary, the two models propose a conceptualization of perception in L2 that is strongly mediated by the available categories of the L1.
In my personal opinion, I see analogies between the concept of perceptual warping and the concept of assimilation. In all the proposed theories there is the idea that speech functions at the basis of a starting mechanism are able to generalize a constant property across several instances of noisy, unsystematic, highly variable of sound patterns. A mechanism that categorizes. Speech sounds are connected with the anatomical configuration of the vocal tract and therefore there has to a be a possibility to map sound instances with the motor activation and motor configuration of the phono-articulatory apparatus in order to process the incoming stimuli as speech (Liberman et al., 1961; Galantucci et al.2006, Best, 1995; Best & Tyler, 2007; Hickock & Poeppel, 2007). In third instance, there is the idea that experience shapes what the perceptual system considers relevant in order to process speech successfully, and the attunement process is designed to discard irrelevant and unsystematic details and to attend what is considered recognizable selectively.
This highlights that attention plays a determinant role in speech perception, once a system is stably represented.
How attention seeks the expected patterns in the incoming stimulus impacts directly on the linguistic processing of both auditive and motoric features that in the course of life-experience were not encountered. It leads in most extreme cases to completely disattend relevant features of the L2, and in a more moderate perspective to somewhat categorize the stimuli adopting a different set of parameters when compared to native speakers of the same language.

1.5.3. Aspects of L2 acquisition of suprasegmental features

The intonational systems of several languages have begun to be extensively studied in the past years (Hirst & Di Cristo, 1998; Ladd, 1996), and nowadays it is possible to instantiate a comparison between different intonational systems and it is also possible to address classical terms (e.g., contrastive analysis) the issues of SLA of intonation. Nonetheless, as it has been described before, the modelling of a theory of SLA of intonation and its processing it is rather hard, given its multidimensional nature, the available theories mainly designed to account for L2 acquisition of segments, and a plurality of experiments that in the years has concentrated on investigating mainly L2 production.

In a comprehensive review of the cross-linguistic research on intonation in the SLA field of 2007, Mennen (2007) and Gut (2007) reported as few as about 20 studies in 25 years specifically addressing the acquisition of intonation in L2. Out of about 20 studies on intonation only 4 have been found addressing the perception of intonation (e.g. Liang & van Heuven, 2007; Holly J. Nibert, 2006). The framework, also, is the one described for general L2 acquisition, thus, considering the analysis of errors and the contrastive analysis of L1 and L2 structures as a primary source of predictions for the behavior of L2 speakers. Further on, Gut (2007) highlighted how most of the studies focused on the realization of specific non-native intonational structure, but no investigation was available addressing the relationship between prosodic domains. In final instance, up to 2007, no study specifically addressed external covariates assuming to play a role in the acquisition of prosody in L2, such as individual variability.

Within the studies focusing on production Mennen (2007) reported the following phenomena, as the most characterizing features of L2 intonation produced by native speakers of several languages, all acquiring English (Mennen, 2007; p.55):

- "a narrower pitch range (Backman, 1979; Jenner, 1976; Willems, 1982);
- problems with the correct placement of prominence (Backman 1979; Jenner 1976);
- replacement of rises with falls and vice versa (Adams & Munro, 1978; Backman, 1979; Jenner, 1976; Lepetit, 1989; Willems, 1982);
- incorrect pitch on unstressed syllables (Backman 1979; McGory 1997; Willems 1982);
- difference in final pitch rise (Backman 1979; Willems 1982);
- starting pitch too low (Backman 1979; Willems 1982);
- problems with reset from low level to mid-level after a boundary (Willems 1982);
- a smaller declination rate (Willems 1982);"

Following the introduction of the AM model in the field of SLA, the analysis began to adopt the separation of the phonological and phonetic level as the main assumption for the acquisitional perspective.

As previously mentioned, Ladd (1996) formulated the 4 dimensions on which systemic variation can occur in a linguo-specific manner (semantic, systemic, realizational and phonotactic) and this overall has led to the identification of L2 production deviating in each specific dimension.

The phonological influence of the native language, specifically addressing the phonological inventory of PAs and BTs in a linguistic system, and several studies have reported the substitution in the target language of rising PA with falling PA and vice versa. (e.g. (Adams & Munro, 1978; Backman, 1979; Jenner, 1976; Lepetit, 1989; Willems, 1982). In addition, phonetic deviations have been found in L2 productions at unusual pitch ranges (Mennen, 2007) steeper rising contour than the ones available in the target language (Ueyama, 1997) and in the alignment of the peak within the target syllable (Mennen, 2004).

Alignment is one of the phonetic aspects of intonation that has been shown it can vary from variety to variety of the same language, as in Swedish (Bruce & Gårding, 1978) and Danish dialects (Grønnum, 1994) and within British English varieties (Grabe, Post, Nolan, & Farrar, 2000), and German dialects (Atterer & Ladd, 2004).
A difference in alignment in production leads to an effect described in Ladd (1996, p. 128) in which both German English and Italian display an H+L* accent in their inventory, but the peak alignment strategy differs across languages, occurring early in Italian, and rather late in English and German. The production of the word "Mantova" in Italian, thus, which presupposes an early peak on the first syllable, is altered in L2 by the alignment of the peak later in the syllable. This "false" alignment perceptually can result in a categorical perceptual change of the PA aligned to the target syllable, since the unusual position of the peak can lead to a false identification of a rise on the target syllable (L*+H).

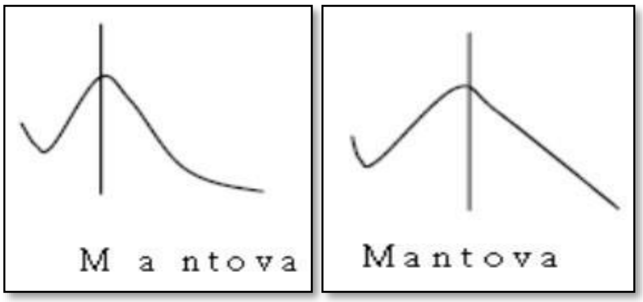


Figure 1.5.1 L1 vs. L2 alignment affecting the position of the perceived word-stress - adapted from Mennen (2007)

Mennen finds a similar phenomenon (2004) investigating the L2 production of the L+H* accent in advanced L2 learners of Greek, native speakers of Dutch, a language that also displays the same PA.
Nonetheless, the accent is aligned to the onset of the post-tonic syllable in Greek (late peak), and it is centrally aligned in the middle of the vowel in Dutch. The results of Mennen (2004) indicated that the alignment strategy of the Dutch is transferred in the production of Greek as L2.
Interestingly, Mennen (2004) reported the presence of individual variability, highlighting a very small proportion of participants succeed in the correct alignment of the peak as required by the Greek target system.

The use of pitch range has been found to vary from language to language. In Ladd (1996), the cross-linguistic differences are described in terms of variation of the level (pitch height) and variation of span (the range of frequencies). An anectodical different use of pitch range is known to exist, for example, between English and German. It appears that English pitch variation has a greater excursion within an utterance when compared to pitch modulation in German (Eckert & Laver, 1994; Gibbon, 1998; Patterson, 2000). The application of the L1 values to speech in L2 results therefore in an exaggerate pitch modulation in German L2, produced by English speakers, and in the impression of a too flat contour in L2 speakers of English, native speakers of German (Trim, 1988).

L2 acquisition appears to be generally connotated by a wrong positioning of the nuclear PA when compared to the metrical and pragmatic rules of the L2 target system. This is one of the cases that is addressed in perception in Chapter 4.

The literature, nonetheless, is mainly concerned again with realizational aspects than structural differences in the construction of discourse. Prominence can be differently realized across languages, adopting different combinations of phonetic features. For example, prominence is expressed through F0 modulation, duration, intensity and spectral information in English (Beckman, 1986), strongly relying only on F0 modulation in a language like Japanese. The L2 acquisition implies, thus, to acquire to modulate more than one phonetic parameter, and implies that differences in stress realization across languages can be strongly related to the phonetic aspects of the language rather than phonological ones (Ling & Grabe, 1999).

1.5.4. **Studies on the perception of L2 intonation**

There are very few studies that focused on the perception of L2 intonation at sentence-level.
One of the first studies, Willems (1982), had the aim to provide a perceptual description of the English intonation to Dutch students, learners of English, and focused on aspects of pitch range in a time of transition between the theories preceding the AM approach (e.g. British School) and the introduction of the metric approach.
(Cruz-Ferreira, 1987) investigated with a crossed design the perception of two intonational contours (e.g., *Didn't John en`joy it* vs. *Didn't John enˏjoy it*), both in English and European Portuguese, by native speakers of the two languages learning the other as L2. She adopted a sort of CP paradigm, and participants had first to state whether the pair of stimuli were the identical or different and then they had to match the two categories representing the two types of sentences with the two stimuli, in their L1 and L2. The results highlighted a main effect of the intonational contours of the native language guiding the pattern of responses in L2. Interestingly, Cruz-Ferreira found a sensitivity of L2 perceivers towards acoustically salient peaks, that in her opinion was used as a pitch height strategy for the identification macro-configurations such as a clearly rising vs. clearly falling contour, not further analyzed in clear cut discrete phonological categories. In final instance, she found an effect of markedness, since participants preferred the less marked interpretation of the contours on the basis of their native language system.
In a study adopting the British School of intonation analysis (Cruttenden, 1997; Gussenhoven, 1984; O'Connor & Arnold, 1973), Grabe, Rosner, Garcia-Albea, & Zhou (2003) tested the perception of 11 possible contours attested in Southern British English. She tested 3 groups of participants: native speakers, and Chinese and Spanish learners of English, having a low proficiency in English L2. The 11 contours can be grouped into two major categories of rising and falling contours. In a first experiment, participants heard a pair of different stimuli and had to rate their similarity on a scale from 1 to 10. In a second experiment, the same procedure was adopted testing a different sample of participants native speaker of the same languages with non-speech material, in which the pitch contour was modulated in the same way. The authors found no significantly different pattern of results in the perception of speech and non-speech material. In addition, all speakers seemed to be able to correctly distinguish rising vs. falling contours adopting, in their opinion, a universal auditory mechanism. The impact of the native language emerged in the setting of the boundaries between categories. Chinese perceivers displayed a much similar correlation in the discrimination strategies between speech and non-speech material, according to the authors they seemed to have adopted for both types of stimuli a non-linguistic strategy.
Nibert (2005) and Nibert (2006) investigated the perception of phrasing through 3 possible prosodic configurations in Spanish (p.133):

1. [[lilas]$_{H-}$ [y lirios amarillos]]$_{L-L\%}$ → 'lilacs and yellow irises'
2. [[lilas y lirios]$_{H-}$ [amarillos]]$_{L-L\%}$ → 'yellow lilacs and yellow irises'

3. [[lilas y lirios amarillos]]$_{L-L\%}$ - → both interpretations are possible

Participants were English learners of Spanish, grouped in 3 proficiency levels, and were compared with results collected from native speakers in a previous experiment (Nibert, 2000). Participants perceived paired of stimuli of distinctive configurations and were required to associate the correct interpretation of the content on the basis of the pitch contour. The two studies highlight an effect of proficiency which increases the sharpness of the perceptual boundaries of the categories: beginners' association of contours and meaning is less restrictive than the intermediate's criterion, that again is less restrictive when compared to advanced learners. The results are interpreted as a gradual restructuring of the interlanguage on the basis of Universal Grammar principles, that at the initial stages is strongly guided by the presence of the native language categories.

Gili Fivela (2012) investigated the perception of non-focal (NF) and corrective focus (CF) construction in English L2, by native speakers of a southern variety of Italian (Apulia, Lecce). Phonetically, in English, the two contours mainly differ in pitch height, on the contrary, in the adopted variety of Italian, the two configurations vary for both the relative pitch height and alignment. Moreover, the two English accents can be classified as belonging to one category in the adopted variety since they both resemble the CF accent. The stimuli were created recording both native speakers of Italian and English producing the target contours, and then two batteries of stimuli were created: one of the native language items, and one where the intonational contour extracted from English L1 speech was superimposed on L1 segmental material. The stimuli were perceived through a CP paradigm and, also, reaction times have also been collected. Results show that L2 perceivers were able to categorize the stimuli, but the discrimination rate was lower for stimuli where the L2 prosody was superimposed, in particular, discrimination rates were lower in the case of non-focal structures uttered with English prosody. Reaction Times congruently were higher for the same category.

Zárate-Sández, (2015) tested the presence of CP in the perception of alignment of the prenuclear accents' positions and final rising BTs in different groups of L2 Spanish learners organized per proficiency (low, high, very high), and native speakers, bilinguals, and monolingual speakers of English without any competence in Spanish. All the pitch contours were aligned on the sentence *La nena lloraba ('the girl was crying').* In this study, the original CP paradigm, composed of the identification and discrimination task, was substituted with an imitation procedure (Dilley & Brown, 2007; Dilley, 2005; Gussenhoven, 1999, 2006; Pierrehumbert & Steele, 1989; Redi, 2003). The participants perceived batteries of stimuli with prenuclear peaks aligned in 10 positions within the tonic syllable and were required to imitate the contour after each trial. The same task was adopted for the final boundary tones rising in 10 steps from low to high.

The data in the production of each speaker and each group of speakers was then aggregated for each level, both in the alignment task and in the rising pitch task. The analysis investigated whether in production the speakers showed to reproduce the stimuli according to the two categories: early vs. late peak and low s. high pitch contour in the final boundary tone, therefore displaying to align the first half of the levels in the first task averagely earlier than the second half of levels, and to reproduce half of the boundary tones as averagely low, and the second half of them as averagely high, in a dichotomic way. The presence of the categories in the production results was interpreted as a cue for the presence of categorical perception. From the original design, the author decided not to consider the data of L2 learners with low proficiency since their production was very discontinuous and strongly influenced by the L1. The results showed the presence of CP in the alignment task and a continuous perception of the final rising tone.

1.5.5. The L2 Intonation Learning theory

On the basis of these preliminary findings, intonational features are recognized as relevant aspects of the L2 acquisition. The L2 production deviating from the native language ones in the realization of suprasegmental features contribute to the perception of foreign-accented speech in L2 learners and intonation is nowadays undoubtedly recognized to be subjected to the process of acquisition as all the other language-structures (Anderson-Hsieh, Johnson, & Koehler, 1992; Jilka, 2000; Munro & Derwing, 1995; Trofimovich & Baker, 2006). In recent years, research has moved forward defining an emerging model of acquisition specific for intonation. I address here the proposal of the L2 Intonation Learning theory (LILt)(Mennen, 2015).

The theory considers, and draws upon, the above-presented issues on the nature of prosodic categories (Gussenhoven, 2006; Ladd, 1996); it distinguishes the role of phonetic and phonological factors by adopting AM as referring theoretical model (Mennen, 1999, 2004, 2007; Pierrehumbert, 1980; Strange, 2007) and aims to verify how well the presented models of L2 perception developed for segments can suit the L2 perception and acquisition of prosody (Flege, 1995; Best & Tyler, 2007).

The LILt adopts from Ladd (1996) the 4 dimensions through which intonation can be investigated cross-linguistically, and it updates them in the following manner (Mennen, 2015; p. 173):

1. "the inventory and distribution of categorical phonological elements (systemic dimension)
2. The phonetic implementation of these categorical elements (realizational dimension)
3. The functionality of the categorical elements or tunes (semantic dimension)
4. The frequency of use of the categorical elements (frequency dimension)"

Analogously to the model presented by Ladd (1996), the systemic dimension considers the structural differences between two languages, and by adopting the AM theory, refers to the phonological inventory of pitch accents, boundary tones, and phrasing phenomena. Typological differences in the use of specific PAs have been found in the comparison for example of English received pronunciation with other varieties, which present combinations of accents not available in Standard English (Cruttenden, 1986), and in the comparison of the possible PAs available in English respect to French, which seems to allow less variation in the choice of intonational contours (Post, D'Imperio, & Gussenhoven, 2007).

The phonetic implementation primarily deals with the cross-linguistical comparison of alignment rules: e.g., nuclear accents are realized later in Dutch than in English (Schepman, Lickley, & Ladd, 2006); prenuclear accents are aligned earlier in English than in German (Atterer & Ladd, 2005); later in Greek than in Dutch (Mennen, 1999; 2004).

The semantic dimension is the dimension that will be specifically addressed in the third study in this work, and it deals with the structural configuration of tunes that express a semantic/pragmatic meaning. For example, the focus is signaled through a specific pitch configuration in European Portuguese (Frota, 2014); yes/no questions are realized with descending in contours in Greek (Arvaniti, Ladd, & Mennen 2006) and in some varieties spoken in southern Italy (Grice, 2017; Grice & Savino, 1995); the modulation of accentuation in function of focus/background structures and given/new information in discourse, shows the presence of typological differences across languages and this will be topic of the EEG study, in chapter 4.

The last dimension is newly introduced by the LILt theory and refers to differences across languages in the use and distributions of specific contours. A language can display the same phonological inventory of another one, but the speakers can systematically and more frequently use a specific contour; the same accent can be less frequently adopted in the other linguist system (Grabe, Kochanski, & Coleman, 2005; Mennen, Schaeffler, & Docherty, 2012).

The difficulties in the acquisition of an L2 system can occur spanning over more than one dimension. A language can diverge in the available PAs but also in the systematic use of specific phonetic features, displaying a preference for early or late alignment, display different proportions in the frequency of use of specific PAs and, and in the ways through which intonation is integrated into discourse with non-marked and marked syntactical structures that are reflected in the phonological tune.

In addition, factors belonging to different dimensions can interact; as in the previously described case, where differences in alignment, or in the realizational steepness of a peak, can result in the perception of a different category. Therefore, deviations occurring at the phonetic levels are misinterpreted as reflecting a different organization of the prosodic categories at the phonological level.

Another aim of the LILt, probably the main aim, is to be able to predict the relative difficulty of acquisition of specific patterns, given a description of the phonological categories, phonetic implementations, and semantic configuration of the languages that are compared. The aim is therefore to predict a pattern of acquisition given the organization of prosodic categories in the mind of the learner given his/her native language and the target one. To address the topic, the theory tries to account for what has been previously established for the perception of segments and tries to implement in the field of intonation.

It addresses firstly, the role of the perceptual "filter" of the L1, which implies the recognition of the notion of category and the fact that perceptually assimilated features will fall within the L1 categories, and highly deviating phonetic features mismatching with the category's prototype of the L1 will trigger the mechanism for a new category formation (Best & Tyler, 2007).

A starting question is raised on the actual ability of L2 perceivers to sufficiently attend to intonational cues in order to perform a perceptual mismatch with the existing L1 categories. Mennen (2015) raises a question on the basis of the presented perceptual studies and of some specific studies addressing lexical tones perception, that I have not discussed in the previous section, that show poor discrimination in function of proficiency in L2 (Gili Fivela, 2012; Liang & van Heuven, 2007; Nibert, 2005, 2006; Trimble, 2013).

The second question is based on the nature of intonation requiring the semantic dimension to identify an intonational contour.

Mennen (2015) and Gili Fivela (2012) propose therefore that explicit reference to the semantic dimension has to be made in order to instantiate the processing of perceptual similarity.

In second instance, the LILt recognizes the phonetic level as crucial in order to clearly understand the way the categories of the L2 are perceived and acquired since they carry fine-grained information. By recognizing the relevance of the phonetic implementation, it also recognized that the phonetic context in which a contour occurs could impact on the discrimination sensitivity of learners.

The third assumption is the recognition of Age of Acquisition (AoA) as a predictor to successful L2 acquisition of prosodic patterns. Also, it is further proposed that age impacts differently in different aspects of intonation. The proposal is made observing some specific evidence in L2 production, comparing beginning, intermediate and advance learners, as in Jun & Oh (2000) where advanced learners of Koreans were correctly reproducing contours on the basis of the systemic level, e.g., correctly marking phrase boundaries, but still showing deviating patterns in the realizational dimension. In addition, despite correctly producing boundary tones they displayed difficulties in mapping the correct semantic function of tones, like the distinctive marking of interrogative yes/no questions and wh-questions.

In final instance, the LILt postulates that a target-like acquisition of L2 prosodic categories remains possible also outside infancy and the critical period since there is evidence of late learners achieving a high level of proficiency in L2 also in the realization of phonetic features.

The presence of exceptional learners will be addressed in the following section. It is relevant here to mention that the LILt is broadly open in postulating that a target-like acquisition is possible in a more distributed manner, and not restricted only to the presence of exceptionally high language aptitude. The exact realization of phonetic features in L2 is found in a low proportion of participants in Mennen (2004) and De Leeuw et al. (2012) and in Bongaerts et al. (1997) at the segmental level; and quite often L2 studies report the presence of exceptionally poor or talented participants performing in the tasks. In a more recent study, Mennen (2015) found a group effect of advanced L2 learners of English, L1 speakers of German, who averagely correctly adopted target-like values of pitch range.

The possibility of a bidirectional influence of both the native and non-native system is exploratively considered, but the evidence available is not sufficient at the moment to make precise predictions. The hypothesis comes from both the SLM and the PAM/L2 both assuming that L1 and L2 coexist in the same phonological space. In studies of intonation, evidence from a bidirectional influence come from merging phenomena, where the phonetic implementation of alignment and the rising of the contour can assume values belonging to both the categories in each language (De Leeuw, Mennen, & Scobbie, 2012). A second phenomenon is "overshooting" that implies that features belonging to the L2, are reproduced in an extreme manner in the L1, obtaining a production that is both deviating from native and second language norms. Instances of overshooting in intonation are found by De Leeuw et al. (2012) where L2 learners of English produce extremely late peaks in L1, later than what the norm of alignment in German would predict.

1.6. Covariates, Individual differences & Language Aptitude

The constant presence of participants in SLA performing exceptionally well, sometimes in a native-like manner, in L2 despite having started the acquisition later in life, and therefore representing a strong exception to the Critical Period Hypothesis (Abrahamsson & Hyltenstam, 2008; Birdsong, 2007; Bongaerts, 1999; Colantoni & Steele, 2006; Reiterer et al., 2011; van Boxtel et al., 2005) has contribute to the establishment of two assumptions that began guiding research in this field. On one side, the idea that some individuals might be connotated by the presence of the Language Aptitude factor (Obler & Fein, 1988; Skehan, 2002), that is able to overcome the constraints of age, its correlated perceptual (in)flexibility, and the difficulties generated by the acquisition of more marked L2 structure, when compared to the L1. On the other side, in a more general perspective, that performance in language-related tasks, in L1 and L2, are explained by both external factors related to exposure to a specific linguistic system, but also by internal factors characterizing the individual on a broader cognitive spectrum comprehending cognitive abilities and personality traits.

Language Aptitude is grounded on the work of J.B. Carroll (Carroll & Sapon, 1959; Carroll, 1993) who first defined three main assumptions, establishing a cognitive framework of investigation, namely: that the cognitive abilities related to L2 aptitude have to be distinctly investigated from other cognitive abilities, such as intelligence; that aptitude is persistent and stable in the individuals; and that the general concept is composed of specific components related to language areas. Carroll's original formulation comprehended 4 levels:

- the phonemic coding ability (of unfamiliar auditory material);
- the inductive language learning ability (to extrapolate and generalize information from the input);
- the grammatical sensitivity (identifying the morpho-syntactic function of words);

- the associative learning ability (to establish a link between available L1 representations and L2 ones).

In the current language aptitude literature, the combination of these skills often emerges in polarized one, generally showing two main linguistic talents. On one side there are individuals that, as predicted, keep a strong native language foreign accent in L2, while for example achieving native-like performances in other language subfields, such as syntax, morphology, and vocabulary. This phenomenon is often referred to as the "Joseph Conrad" effect, referring to the famous writer, who was a native speaker of Polish, and acquired English as third L2, and who undoubtedly achieved in his writings in L2 English a native-like competence, but, on the contrary, maintained a strong Foreign Accent in his speech.
Individuals who display, e.g. the Joseph Conrad effects are considered individuals with "a talent for grammar" (Nauchi & Sakai, 2009), individuals acquiring native-like phonological structures and phonetic implementations are referred as individuals with “a talent for accent” (Reiterer et al., 2011).
Language Aptitude nowadays considers a multifactorial source of variation that comprehends the role of: short-term, working, and long-term memory (Gathercole, Pickering, Knight, & Stegmann, 2004; Wen, Mota, & McNeill, 2015; Williams, 2013); attention (Robinson, Mackey, Gass, & Schmidt, 2012); types of intelligence and mental flexibility (Martin-Rhee & Bialystok, 2008); it considers also personality traits, such as introversion and extroversion; the managing of anxiety and distress (Dörnyei, 2006); and the presence of constructs evaluating the ability to assume other individuals’ perspectives such as the Theory of Mind concept, and the empathetic capacities (Guiora, Brannon, & Dull, 1972; Rota & Reiterer, 2009). In recent years, the neuroscientific framework, also began to seek for neural patterns explicitly related to the presence of specific language talents such as the imitation ability in auditive tasks, previously described in the section where the studies on perception L2 intonation were presented (Amunts, Schleicher, & Zilles, 2004; Golestani & Zatorre, 2009; Reiterer et al., 2011).
Roughly the plurality of internal factors comprehends two major areas impacting on L2 acquisition: factors directly pointing to cognitive aspects such as memory, attention and mental flexibility, and factors concerning the personality type and traits that strongly relate to the use of language within social interactions and can interact with strictly cognitive mechanism during the L2 processing.

The literature on individual variability on second language learning separates the acquisition in a natural context, among L2 native speakers without formal instruction, and the learning occurring in class-settings through guided instructions. This determines a differentiation of the patterns of acquisition that can occur through implicit or explicit mechanisms and can create two types of mental representations.
A learner can have a declarative knowledge of the structures of the language he/she is learning (e.g. knowing the phonological inventory of the L2 is composed of a higher of vowels than the ones available in the L1), or on the contrary can store and access new mental representations in a form of a procedural knowledge (e.g. possess the ability to detect and reproduce vowels that are not present in the L1).

> “Declarative representations are objects of thought, whereas procedural representations provide the (cognitive) actions to work upon these objects” (Gade, Druey, Souza, & Oberauer, 2014, p. 174).

Declarative and procedural knowledge of a phenomenon overlap with the concepts of explicit and implicit knowledge. Explicit knowledge implies the concept of awareness (Paradis, 2009; Ullman, 2015). Formal instruction in class provides explicit and declarative knowledge of the linguistic structures that through practice acquire a procedural representation and are further automatized (DeKeyser, 2017).

During formal instruction learners can acquire, then, the new linguistic structures through explicit learning if they are made aware of the structures to which will be exposed, and the rules with which they combined and organized in the L2 grammar; they can acquire through implicit learning, meaning that they come to know certain structures without being explicitly aware of the rules governing them; and they can acquire them through incidental learning where participants performing a task on a different domain (e.g., semantics) become aware of rules governing related phenomena (e.g., intonational features).

Some authors support the view that individual variability plays a role only in explicit and conscious learning (S. Krashen, 1994, 1999; Reber, 1993). According to Krashen and other authors (e.g., Schwartz & Sprouse, 1994) the mechanisms of acquisition of the L2 share the properties of acquisition of the L1, therefore given an innate ability to acquire language across individuals, unconscious and implicit learning should affect equally the population and be explainable in terms of Universal Grammar mechanisms. This position is debated and also questioned by neurophysiological results highlighting different neural networks responsible for explicit and incidental learning (Paradis, 1994).

The role of attention is raised several times in the previously presented literature. Its role is recognized within the abstract notion of category (Robert, 2005), it is at the base of the mechanisms modeled by the SLM and the PAM/L2 (Flege, 1995; Best and Tyler, 2007), it is the mechanism that grants the identification of an established category since the identification process requires to selectively search for a recognizable feature or pattern (Harding & Cooke, 2007; Kalinli & Narayanan, 2007), and it is at the base of the explanation furnished for a cognitive interpretation of CP of alignment (Niebuhr and Kohler, 2004), in final instance it is among the reasons that lead the LILt model to discuss the notion of prosodic category in L1 and L2, given the evidence that intonation is often poorly discriminated in L2 learners (Mennen, 2015).

Since, in chapter 4 the perception of specific intonational categories in German is observed in a group of participants in a condition of immersion, and sharing a level of proficiency ranging from intermediate to high, but that nonetheless acquired the language along the years in a heterogeneous manner (some participants have undergone many years of instruction, some participants have been intensively exposed to native speech) I discuss here how the mechanism of attention is reported to relate to acquisition, in order to possibly observe whether different educational patterns can enhance this mechanism or not.

1.6.1. Cognitive factors

Analogously to the literature proposed by scholars specifically investigating perception, category formation, the perception of category in L1 and L2, the perception of intonational categories in L1 and L2, the literature on language aptitude in second language acquisition proposed models to individuate the relevant cognitive abilities and personality traits that can have an impact on the process of acquisition.

Interestingly despite the different aims of each branch of research, the models somewhat converge, in this case too, in identifying the role discrimination and generalization, of identification of recognizable features and their organization and restructuring.

Skehan (2013, pp. 385-386), in line with the presented literature, proposed a set of relevant stages in L2 processing, that deal, above all, with working memory and attention.

1. Input processing and Noticing
2. Pattern identification
3. Complexification, restructuring, and integration

4. Error avoidance
5. Repertoire and salience creation
6. Automatization and Lexicalization

The role of attention has been claimed in each section of the literature that has been previously described. Its role is recognized within the abstract notion of category (Robert, 2005), is at the base of the mechanisms modelled by the SLM and the PAM/L2 (Flege, 1995; Best and Tyler, 2007), is the mechanism that grants the identification of an established category since the identification process requires to selectively search for a recognizable feature or pattern (Harding & Cooke, 2007; Kalinli & Narayanan, 2007), is at the base of the explanation furnished for a cognitive interpretation of CP of alignment (Niebuhr and Kohler, 2004), is among the reasons that lead the LILt model to discuss the notion of prosodic category in L1 and L2, given the evidence that intonation is often poorly discriminated in L2 learners (Mennen, 2015).
The role of attention is also debated in the research pertaining to individual variability in second language acquisition. Some theoretical positions assume that It guides the difference between what is considered the input and what assumes the status of intake (Corder, 1967; Robinson et al., 2012). It is considered input "what is available to be learned" (Robinson et al., 2012; p.248) and it is considered intake what learners cognitively register during perception and make available for further processing. These two concepts are also associated to steps in L2 processing: at the initial stage, perception involves the formation of a "preliminary intake" that in the following stages is recoded and integrated with information available memory and ends with the "final intake" stage (Chaudron, 1985). According to these theoretical positions, a core aspect in the process of acquisition is the step that leads to the “intake” formation from an initial simple input perception. This specific issue is addressed by research on selective attention and processing awareness (S.M. Gass, 1988; Robinson, 1995; Robinson et al., 2012; Schmidt, 1992). Gass (1988) interestingly introduced the concept of "apperception," defined as "the process of understanding which newly observed qualities of an object are related to past experiences," and that leads to the noticing by the learner of some specific L2 features. The noticing of specific features is responsible for the enhance of meta-linguistic reasoning and, thus, the process of L2 acquisition (Schmidt, 1990).
The debate on the role of attention, nonetheless, is kept alive by theoretical positions that assume that acquisition does not necessitate this preliminary perceptual stage, and that “non-attentional” acquisition is possible (Schacter, Rounds, Wright, & Smith, 1996), and that explicit knowledge of a certain phenomenon can be reached in time, even if at the initial stages of acquisition it had been implicitly acquired, thus, without specific awareness (Ellis, & Ellis, 1994, p. 95).

A second aspect that is related to the modulation of attention, and it is accounted as a relevant factor in the perception of intonation, especially impacting on the mechanism of intrinsic vs. extrinsic reference comparisons is the capacity to retain information in working memory (Skehan, 2002). Several models are addressing the functioning of working memory (Baddeley, 1992; Baddeley & Hitch, 1974; Conway, 2005; Engle, 2002; Miyake & Shah, 1999). Its modeling is not the direct object of this work; therefore, it will not be extensively explained here (for a review see Williams, 2012). Nonetheless, I will discuss in the next chapter how the categorization process, independently from the nature of the signal (speech vs. non-speech), relies on the ability to retain information after the perception of the stimulus in order to instantiate the evaluation of the perceived stimulus with the available representations. The representations, in the case of a perceptual experiment, can be both the categories of the native languages, but also, they can be categories that have been perceptually learned in the experimental task if it requires it. This is expressed by a model developed in studies of categorical perception of tonal movements (Xu et al., 2006) and I will adopt it as a possible explanation in the results of experiment one. Since the experiment one requires to possibly instantiate a perceptual learning mechanism on the basis of pitch height, and the task is failed by about one-fifth of the participants, individual differences in working memory could be a relevant factor to take into consideration in future research.

A third aspect that I consider relevant is exploring how individual variabilities that can affect the perception of pitch with linguistic function in L1 and L2 is the sensitivity to pitch variation. When interested in the processing of pitch perception this is in truth the first stage where an individual can vary. Literature in L2 and generally in pitch perception often speaks about outstanding skills of some participants in the study, who are often classified as "pitch listener." On the contrary, the hypothesis that sensitivity to pitch distributes in the population, letting emerge the possibility that some learners could be poor listeners, is often not considered.

Music and language are both taking advantage of a mechanism responsible for the recognition of sound pattern, what in ultimate analysis can be interpreted as “the category” in this work, nonetheless it has been shown that there are several networks responsible for this operation, specialized in different domains(Polster & Rose, 1998). The plurality of networks is deducted by the presence of patients displaying different types and degrees of dissociation and presenting the loss of the ability in one domain that, nonetheless, does not affect other domains (for a review see Polster & Rose, 1998). The study of patients allowed to observe that the mechanisms underlying music and language are not shared and are neuroanatomically distinguishable (Peretz, 2001). In particular, evidence highlighted that the networks responsible for voice recognition and voice discrimination are independent (Van Lacker, Cummings, Kreiman, & Dobkin, 1988).

Congenital disorders, on the other hand, refer (Peretz, 2001, p.156):

> “to unexpected failures or preservation in a particular cognitive domain in comparison to the general level of intellectual and socioemotional functioning. These deficiencies are termed congenital since their presence can be detected very early in development.”

The Autism Spectrum Conditions (ASC) are considered a form of congenital disorder, even if the exact etiology is unknown. Interestingly 1 to 10% of the population diagnosed with ASC displays a remarkable musical sensitivity, displaying on the contrary, atypical processing of language aspects within the auditory domain. This is particularly interesting for the nature of pitch given its multidimensional function. In the last paragraph of this section I report, in fact, that the presence of traits belonging to the spectrum of autism correlates with a less efficient and accurate processing of the pragmatic aspects of communication, which indeed, in synergy with other linguistic dimensions expressed through segmental information, are governed by pitch modulation in language.

The disorder, having congenital origins or subsequent to brain injury, is amusia, or tone-deafness (Patel, 2010; Patel, Foxton, & Griffiths, 2005). Tone-deafness interests those individuals, actually estimated as the 5% of the population, that display values of detecting thresholds of pitch variation extremely high, impairments in the ability to judge the direction of tones (rising or falling), and deficits in the discrimination of tone sequences outside the domain of music (Ayotte, Peretz, & Hyde, 2002; Foxton, Dean, Gee, Peretz, & Griffiths, 2004; Isabelle Peretz & Hyde, 2003). Adopting a specific battery, designed for brain-damaged patients, Peretz (2001) investigated the use of melodic contour and the discrimination of rhythmic patterns of 16 congenital amusic and 60 normo-typical non-musician participants. She found that the normo-typical population did not display any impairment in both the melodic and rhythmic task, where on the contrary

88% of amusical population displayed a deficit in the processing of melodies. The author accounted for the deficit with possible auditory problems impairing pitch discrimination.
Ayotte et al. (2002) found an analogous dissociation of the two systems. Investigating the perception of tonal sequences aligned to speech and in non-speech conditions, the authors found that individuals presenting amusia had substantial deficits in the non-speech condition but could discriminate the intonational sequences aligned to speech material. On the contrary, normotypic individuals displayed to perform equally well in both tasks.
One hypothesis suggested that differences in the processing of intonation with linguistic function and intonation within the musical domain were caused by the fact that relevant intonational movements in language, sufficiently exceed the discriminations' threshold values and therefore are processable also by individuals with congenital disorders (Peretz & Hyde, 2003). Patel et al., (2005) performed a discrimination task with individuals affected by amusia, where for 3 types of stimuli patients had to states whether the presented pairs were identical of different stimuli. The 3 types of stimuli were based on speech material on which intonational contours were modulated, non-speech discrete analogs, and non-speech glide analogs with glides identical to the one occurring in the speech condition. The pitch modulation was the opposition of low vs. high boundary tones identifying statements vs. questions. The results highlighted that the discrimination accuracy was significantly and equally lower in both the 2 non-linguistic conditions and that intonation in language was discriminated with higher accuracy, in line with Ayotte et al. (2002). The authors conclude that reasons for the lower sensitivity in musical decoding in the pathological population have to be found in higher order processing mechanisms.
These results are particularly interesting since one aspect that is known to characterize the perception in a foreign language is the inability to correctly segment the fluent continuous speech signal in word units because the lexicon and the phonological inventory of phonemes and allophones are unknown.
This is known to affect the recognition of language primarily. We describe the speech of speakers of an unknown language often on the basis of an impressionistic auditory sensation. Of the most famous is the description of the way foreign and unknown populations spoke outside the borders of the Greek and Roman territories in ancient times, that lead to the formation of the word "barbaric," literally, people that make the sounds "bar-bar" when they speak.
If the mechanisms underlying pitch processing in speech and music are distinct and independent, it is possible that a lower degree of recognition of segments affects the sensitivity to pitch discrimination with linguistic function, since it is possible that instantiates a processing based only on auditory "musical" properties.
The neurophysiological investigation of tone perception in native speakers of tonal languages and non-native speakers of tonal languages has highlighted that a mechanism of this nature occurs and relies on lateralized networks. This specific literature is discussed in chapter two in the introduction to the first experiment.
It is possible therefore that individual variability is expressed in the capacity to infer a linguistic function also in cases where the recognition of segments is impaired or more effortful because requiring to decode phonemes not available in the native languages, through the mechanism described by the PAM/L2, and because the access to the lexical meaning, to the syntactic structure, and ultimately to the pragmatic interpretation is impaired.

1.6.2. Pragmatic and social factors

The last aspect that I discuss is the second group of factors known to play a role in individual variability in language acquisition.
This group of factors deals with personality aspects involved in social relations. Research on Language Aptitude began to address the role of empathy in L2 acquisition with the work of Guiora (Guiora, 1967; Guiora et al., 1972; Guiora, Lane, & Bosworth, 1967). In Guiora and colleagues' perspective, the acquisition of native language in infancy takes place in a developmental stage characterized by ego-permeability, that decreases with age. The concept of language-ego comprehends the individual's self-representation that becomes stable in adulthood and could explain why some individuals are reluctant in losing their native accent and resist to fully acquire an L2 system defending the integrity of their self-representation and identity. Within this framework, L2 acquisition takes the shape of a new identity formation. Empathic traits in the personality, therefore, are relevant variables promoting the ego-permeability.

Another line of research, originally coming from outside of the SLA field, on the other hand, addressed the concept of empathy and recognition of other individuals' mental states – Theory of Mind (ToM) (Baron-Cohen, 1997; Baron-Cohen, Tager-Flusberg, & Cohen, 1994; Frith & Frith, 2006; Frith & Frith, 2005; Leslie, 1991; Surian, 1996) investigating atypical development and the population who suffered from brain injuries in areas involved in the computation of language.
ToM refers to the cognitive capacity to attribute mental states to self and others. It includes the ability to infer others 'intentions, desires, and knowledge, and it represents an essential and fundamental social skill, granting successful interactions, mutual exchange and cooperation among individuals (Ahmed & Miller, 2011; Bradford, Jentzsch, & Gomez, 2015; Hamilton, 2009). The ToM construct is known to typically develop in children between the age of 4 and 9, and it has been mainly tested through the false-belief paradigm (Onishi & Baillargeon, 2005). The paradigm mainly consists in presenting to the child a scene where an experimenter comes into a room, and leaves an object into a box, and then leaves the room. A second experimenter comes into the room and changes the position of the object, moving it in another box. The task requires the child to indicate where the former experimenter will look for the object, once he comes back into the room and he is not aware that the object has been moved from the original position. To correctly indicate where the first experimenter will look, the child has to disentangle what he/she personally knows, having seen the whole scene, and what the first experimenter knows, not having seen the second person coming into the room. The child, in this case, has to take the perspective of another individual (the first experimenter) in order to answer to the task correctly.
The ability to assume other individuals' perspectives then plays a significant role in many social aspects of life (Rosenthal, 1977) and also in language, especially in all aspects guided from a pragmatic function of communication.

This is relevant in this study since it is known that a major role in the modulation of intonation is played by the pragmatic rules of communication, such as the discourse information structure (Chafe, 1974, 1994).
In fact, there is coherent evidence coming from different fields highlighting that individuals who suffered a brain injury in areas connected with the decoding of pragmatic aspects of language, and children displaying developmental disorders as the ones belonging to the Autism Spectrum, display processing differences and impairments in the decoding of pragmatic aspects of language (Siegal & Peterson, 2008; Surian & Siegal, 2008). Behaviorally the processing impairments involve the ability of perspective taking, the auditive perceptual processing, and/or during production, the modulation of pitch conveying both linguistic,

pragmatic functions. The affected aspects comprehend the correct handling of turn-taking during the conversation, the expression of sentence modality, and the paralinguistic functions such as the expression of emotional states. Focusing on the individuals with within the Autism Spectrum Conditions (ASC), alongside with the presence of social and communication difficulties, this population is also characterized by unusually narrow interests and stereotyped patterns of behavior, including resistance to change (Ruta, Mazzone, Mazzone, Wheelwright, & Baron-Cohen, 2012). In particular, high functioning children with autism are reported not being capable of keeping joint attention with the interlocutor (Surian & Siegal, 2008) to engage less than neurotypical individuals in social mutual gaze (Leekam, Hunnisett, & Moore, 1998; Leekam, López, & Moore, 2000); to show verbal deficits in the correct use of third persons' pronouns and deictics (Frith, 2003) and in final instance they display difficulties in the decoding of non-literate meaning as in ironic and metaphoric speech (Eales, 1993; Happé, 1995; Wang, Lee, Sigman, & Dapretto, 2006; Ziatas, Durkin, & Pratt, 2003). Specifically, pertaining pitch, even if a clear picture has not emerged yet, individuals with ASC are reported showing an unusual use of prosodic cues and difficulties in comprehension (Nadig & Shaw, 2012; Rutherford, Baron-Cohen, & Wheelwright, 2002; Shriberg et al., 2001). In what direction, and what precisely the term "unusual prosody" refers to, is still controversial. In fact, some early studies reported the presence of "flat" intonation (Surian & Siegal, 1999), while others reported an overmodulation of pitch (Diehl, Watson, Bennetto, McDonough, & Gunlogson, 2009; Edelson, Grossman, & Tager-Flusberg, 2007; Fosnot & Jun, 1999; H. Green & Tobin, 2009; Nadig & Shaw, 2012; Sharda et al., 2010) reflecting in an "exaggerated intonation". Nonetheless, it is emerging that the modulation of pitch has a clear relation with the development of pragmatic features of language and a growing number of studies has begun to investigate pitch modulation in early infancy, considering anomalous pitch modulations in baby-cry a relevant cue to consider and to observe that can favor early ASC Diagnosis (LaGasse, Neal, & Lester, 2005; Sheinkopf, Iverson, Rinaldi, & Lester, 2012; Sirviö & Michelsson, 1976; Wasz-Hockert, 1968; Werner, Dawson, Osterling, & Dinno, 2000).

Literature on language talent on one side, and literature on language impairment on the other one, can help, therefore, to disentangle which ones among the several factors pertaining specific cognitive mechanisms, such as the attentional processing and the access to memory, and personality traits and perspective taking abilities, such as extroversion, anxiety, empathic concerns and the developing of ToM, impact the most the L2 acquisition of suprasegmental features. The analysis of these individual factors, on the other hand, can allow the developing of tools that by investigating prosodic processing can highlight the presence of related impairments and improve and help the early diagnosis of developmental disorders.

1.6.3. The adopted Questionnaires as correlated measures

In each study of this thesis, I have administered the Autism Spectrum Quotient (AQ) (Baron-Cohen et al., 2001) and the Interpersonal Reactivity Index (IRI) (Davis, 1983, Albiero et al.,2006). The AQ is composed of 5 subscales measuring difficulties in relational abilities, imagination and atypical attentional-(dis)engagements, connected with the presence of autistic traits in the spectrum. The 5 subscales are: Communication (Comunic), Social Abilities (Ab_Soc), Imagination (Imag), Local details (Att_Dett), and Attention Switching (Sp_Att). Each scale is composed of 10 items. Each item can be answered with a 4 levels Likert Scale, ranging from: "strongly disagree," "partially disagree," "partially agree," "strongly agree." The original scoring system does not consider all 4 levels but assigns, depending on the direction of the item, 0 to the pair of levels disagreeing with the item, and 1 with the pair of levels agreeing with the item. In items with the reversed direction of effect, the scoring is reversed too. This generates a roof-score of 10 for each subscale and a total value of 50 for the overall scoring of the AQ. Neurotypical population displays total values in the range of 0-32 points, with an average score of 16. Literature highlights gender differences with men scoring higher than women.

Since I have administered the test as a correlated measure of individual variability and not for clinical reasons (validation data in order to make a diagnosis are not needed), I adopt here the scoring system considering all 4 levels, with a Likert Scale assuming values from 0 to 3. The scoring is reversed for reversed items. This sets a roof-values of 30 in each subscale and a total score of 150 for the overall AQ measures.

The IRI is composed of 4 subscales measuring overall the Empathy construct in two main dimensions: the emotional and the cognitive one. The 4 subscales measure the attitude of the participant to fantasize, to assume interlocutors' perspectives in communication, to show emphatic concern, and to display sign of distress in critical situations and they are: Fantasy (FS), Perspective Taking (PT), Emphatic Concern (EC), and Personal Distress (PD). Each scale is composed of 7 items, and the total test of 28. Each item can be answered on 5 levels Likert Scale, scored from 0 to 4. Literature highlights gender differences with women scoring significantly higher in the FS and PD scales, compared to men.

I consider the total amount of 9 subscales independently as a measure of difficulty or easiness in the handling of social situations, depending on the direction of each dimension in each subscale. The aim is to observe whether the categorization process of pitch with a linguistic function and its relationship with the presence or absence of recognizable segmental materials correlates with individual differences in communication abilities.

The prediction is that the presence of social impairments associated with the spectrum of autism correlates (Ab_Soc, and Comunic) with a lower sensitivity to pitch modulation. On the contrary, the ability to assume the interlocutor's perspective (PT) and higher empathic responses (EC) to other individuals' experiences should represent an advantage in tasks involving the decoding of a pragmatic function related to the pitch modulation.

The two scales of the AQ test, explicitly addressing the property to direct an extreme level of attention to details (Att_Dett) and to be resistant to disengagement (Sp_Att) can interestingly relate to pitch processing two-ways. As typical measures of the spectrum they can display a trend in the same direction of the two social scales (Ab_Soc, Comunic) and be associated with a lower sensitivity to pitch modulation in function of pragmatic aspects; on the contrary, given the fact that the task is autonomously performed at a computer and does not involve a real relational aspect of communication, higher levels of attention directed toward one single modulated parameter during the experiment, F0, can enhance the sensitivity. My prediction is therefore that these two subscales can correlate in both direction with the perceptual task, but negatively correlate with PT and EC scales.

The remaining subscales addressing imagination in both tests (FS, Imag) should then negatively correlate between each other. In final instance, the subscale addressing the personal distress (PD) can be a general indicator, independent from traits belonging to the spectrum, of how distress during perceptual processing in communication can affect the auditive evaluation of the stimuli. The literature on the IRI test highlights that PD negatively correlates with PT, meaning that individuals who highly concentrated on their response to critical situation take less into consideration other individuals' positions and reactions. PD, in this way, should then negatively correlate with PT and EC, and positively correlate with Ab_Soc and Comunic.

1.6.4. Three open questions in perception

1.6.4.1. Summary and open questions on the nature of intonation

The Autosegmental Metric Theory of Intonation elegantly divides in a generative approach all the levels of information that co-occur with the presence of an intonational tune, within the domain of an utterance.
The approach allows identifying primitives in intonation which are simple tones assuming only two values of low and high and that can combine usually forming 4 typical bitonal units representing rising and falling contours aligning to the target syllable targeting either the low or the high tone. Further variations in the accents type are granted by the down-step symbolism and the by a rich and extremely detailed description of the phonetic implementation, mainly focused on the role of alignment, pitch height, pitch range, contour shape, intensity, and syllabic duration.
The formalization also allows the comparison of the intonational structure of different languages and therefore represents a progress for the description of prosodic features.
The notion of category derives almost naturally from such a clear description as it is often reported in association with the abstract phonological representation.

> "Pitch can be modulated in a categorical way, with the *presence vs. absence*, or *type* of pitch movement, and in a gradient way, involving e.g. variations in the way a pitch movement is realized: the extent of the rise or fall, or the pitch range within which a pitch movement is realized."(Grice & Baumann, 2007; p.26)

A direct implication of this interpretation is that, if the interest of investigation is, for example, the acquisition of the intonational system of a foreign language, the core aspect of acquisition will rely first in the mapping of the abstract representation and then in the following learning through procedural rehearsal of the numerous shades in which a discrete category can be realized.
This is not the case for at least two reasons:
In order to establish a discrete abstract phonological representation of a PA or a BT, it is not only necessary to map an acoustic pitch contour, but the contour needs to be aligned to segments organized in syllables, granting the recognition of words, the access to meaning, and the further construction of a sentence-meaning. The discreteness of single intonational unit is reached, and declared in the starting assumptions of AM theory, through the presence of multidimensional cues, that are linguistically organized and predictable in the metrical structure, and with which the personal communicative intentions of the speaker overlap in the construction of a discourse model that connects in the concept of "shared information" sequences of communicative turns.
The turns contain semantically and pragmatically coherent information, and intonation reflects this overall discourse coherence.
The second reason why the notion of category and/or of discreteness is complicated to apply to prosodic units is that clear evidence of classical categorical perception is found in a phonetic aspect: peak alignment. Moreover, the information implemented at the phonetic level shows not to be all of the same nature. Boundary Tones are found in the presented literature to be continuously perceived, I will present in chapter 2 literature explicitly addressing the Boundary Tones disambiguating statements and questions, and a plurality of results is available. Some studies also claim categorical perception, and many studies claim quasi-categorical perception.

It appears, thus, the phonetic shape and alignment of a contour do not carry only sociolinguistic and paralinguistic information, such as a specific style of speaking of a subject, but it carries substantial information to integrate the acoustical signal with higher level linguistic structures.

The co-occurring segmental information influences the way in which the psychophysical properties of pitch are perceived. Fast consonant transitions reduce the intelligibility of the pitch contour, the alignment on a vowel increases it. The velocity with which pitch varies aligned on a vowel also impacts on the discrete recognition of a category, faster and higher peaks are more easily individuated when compared to slow and low movements.

The absolute frequency values play a role in the discrimination sensitivity given neuroanatomical constraints of the hearing system, and the connected presence of a critical bandwidth responsible for the different perception of resolved and unresolved harmonics.

The ability to discriminate and generalize of the basis of a multidimensional signal shows a relationship with age, and it decreases in adults compared to children.

The further application of the available perceptual models, developed for the perception of segments, in the field of second language acquisition of intonation implies the adaptation of a rather simpler notion of category to a complex multidimensional signal. Also, it forces to match the assumptions that are developed in the historical development of the perceptual models with the intonational categories.

Namely, they force the presence of categorical perception since the origin of the theories on segment's perception is attributable to the seminal work of Liberman, and further developed in the years in derived and parallel approaches such as the direct realist view. There is a strong claim in the original development of the Motor Theory of Speech Perception and in the Direct Realist approach at the basis of the PAM/L2 that the discrete nature of the segmental representation is granted by the mapping of the acoustic signal with motor representation. The motor representation of the vocal tract in the production of intonational features, in native and second language learning is rarely addressed, despite the fact that it is recognized.

Aspects of speech contributing to intonation		
Perception	Articulation	Acoustics
Pitch perceived scale: High – low	Quasi-periodic vibrations of vocal folds	Fundamental frequency (F0) Measure: Hertz (Hz)
Loudness perceived scale: Loud-soft	Articulatory effort, subglottal air pressure	Intensity Measure: decibel (dB)
Length perceived scale: Long – short	Duration and phasing of speech gestures	Duration of segments Measure: millisecond (ms)
Vowel quality perceived scale: Full - reduced	Vocal tract configuration, Articulatory precision	Spectral quality Measure: formant values in Hz

Table 1.6.1 Relation between perception, motor-source and acoustical properties of the signal - adapted from Grice & Baumann, 2007; p. 27

In further instance, the PAM/L2 model conceives the perception of phonetic aspects as the perception of invariants and considers the phonological representation of a language system, the organized relationships between constellations of invariants that the learner firstly acquires and then seeks in language perception. It is not clear then how the complete separation of the levels of phonological and phonetic implementation within the AM, often now used as referring system to describe the different linguistic systems congruently

match a model of perceptual learning that postulates an almost integrated level of phonetic and phonological representations.
There is, therefore, the necessity to re-discuss how the several cues present in the speech signal are processed and identified as unitary categories, before successfully approaching a modeling of L2 acquisition of prosodic features.
Intonation demands a higher cognitive load compared to the discrimination and identification of segments. The process of matching and mismatching a stimulus with an identifiable category requires to store longer segments of speech and to operate a comparison on the basis of extrinsic reference.
This demands the ability to notice shape and alignment of the contour, to discriminate between minimal phonetic variations in the signal finely and to retain it in memory and evaluate it with the overall information integrated into the discourse at the semantic/syntactic and pragmatic level.
The individual variability to selectively attend relevant cues and to retain it a specific amount of information for more extended periods can affect the mental formation of a discrete unitary representation.
The individual variability in the evaluation of emotional states, social intentions, and pragmatic coherency of the discourse can affect in final instance the final integration of all relevant information at low sensory and high cognitive level.
This raises many questions on how the actual categorization process of intonational features is performed, on what aspects are shared across individual and on what aspects are relevant to better predict successful processing, in further divided subgroups of the population, displaying specific cognitive and communicative profiles. What happens?

- when pitch movements are aligned to segments that diverge from the native language segments as the PAM/L2 describes;
- when the syllabic structure and the phonotactic implementation diverge from the native one
- when the tune-text association is performed with aligning mapping rules that differ from the native one;
- when the syntax follows a word order that differs from the native one impacting on the metrical organization
- when all the previous level mediates the lexical access
- when the ultimate integration of all the levels is organized in models of discourse structure functioning on typological different mechanisms?

1.6.5. The studies

In this work I address 3 questions on the nature of the categorization process involving pitch categories and on the nature of the integration of categories within a discourse model in native and second language acquisition; and I exploratively address whether individual differences attributable to traits related to the presence of AS conditions are correlated with psychoacoustical effects during the perception of pitch.

In particular, I address in experiment 1 (Chapter 2) the role of the recognizability of segments in the identification and categorization process, investigating whether the decrease of linguistic information carried by segments can be modeled with the modulation of the mathematical parameters of a logistic psychophysics curve fitted on identification and categorization data.
In particular I focus on the two distinct perceptual mechanisms that appear to independently allow the processing of pitch with linguistic function and the processing of pitch within a "musical" or more generally non-linguistic function investigating whether individuals with different profiles in the socio-communicative skills equally can mentally reconstruct a linguistic function in the absence of recognizable segments.

In experiment 2 (chapter 3) I investigate in two conditions at the opposite extremes in the first experiment, namely: the linguistic and non-linguistic condition whether Categorical/Gradual perception emerges with analogous effects in presence or absence of speech material and whether it is related with minimal discrimination threshold of pitch movements. The presence of a peak in the discrimination function in classical CP highlights and is interpreted as the higher discrimination rate of the perceiver. In this experiment, I substitute the original level-comparison adopted in classical paradigms, with an adaptive procedure aim to estimate the minimal discrimination threshold of the participant. If the presence of a category boundary enhances the discrimination and the minimum threshold could be reduced in correspondence of the categories turning point.

In experiment 3 (chapter 4), I investigate by means of Electroencephalography (EEG) the perception of 3 pitch contours (1 congruent and 2 violations) and their association to the discourse structures in function of both given/accessible/new discourse status of a referent and in presence of a particularly high pitch modulation inducing a contrastive focus interpretation.
The experiment is carried out with native speakers of German and intermediate to advanced second language learners of German, native speakers of Italian, living in Germany at the time of the experiment.
The adoption of an online measure such as the Event-Related Potentials (ERP) allows to investigate both early attentional mechanisms, specifically related to auditory processing, and the later discourse structure construction composed by the integration of the previously mentioned information (anaphoric linking) and the following updated required by the introduction of a new referent in the discourse. Early auditory processing components emerge in the time-window at about 200 ms, later integration is observable in the modulation of a biphasic wave at 400 ms and between 600 and 800 ms.

In all 3 experiments, individual personality traits are acquired from participants through the use of the specific questionnaires and the pragmatic profile of the participants is modeled with the outcome of the psychophysical tasks.

Chapter 2
The role of segments:
Categorically perceiving vs. Categorizing while perceiving

2.1. Outline

In this chapter, I present an experimental paradigm based on only the identification task within the classical CP paradigm. The experiment is designed to investigate the role of lexical and segmental information in the processing of graded pitch movements. The task requires to classify the modality of a short utterance (statement vs. question) as a function of the rise in sentence-final position (Boundary Tone – BT). This experiment has 3 goals.

The first goal is to set a baseline in the literature on the perception of intonation in Italian since there is not an equivalent experiment in the literature. Being the first experiment, it aims to assess at group level whether the opposition between descending statements and rising polar questions is dichotomically identifiable in Italian by Italian L1 speakers. Having set such a baseline, the experiment aims to observe if the absence of a recognizable meaning in the utterance (using regular Italian pseudowords) and the presence of foreign phones (pseudowords with foreign phones) affects the way the BT is identified with respect to native language (using words). In a controlled psychophysical paradigm, this is what happens when perceiving a foreign language, in the course of acquisition, where the phonemic inventory might be different from the one in the L1, and the exact meaning representation of the content might not be fully accessed. I hypothesize that the reduced "quality" of information in these two levels also affects the identification of the ultimate pragmatic function of pitch and that it is a relevant aspect to consider in the debate on the nature of intonational categories.

The second goal that follows from the previous steps is "how much" linguistic material is necessary to trigger the identification of pitch within speech so that an intonational category is linguistically recognized. Namely, whether the utterance modality remains available also in the absence of segments but in the presence of a vocal signal (using a further humming condition). According to the discussions outlined in Chapter 1 (paragraph 1.4, and paragraph 1.6.1), the perception of pitch can switch from being interpreted as speech parameter, to being perceived as a modulated non-speech signal. The likelihood of this change is experimentally addressed by introducing a further experimental factor: the order of presentation of the stimuli (from speech to non-speech and vice versa). The driving idea in setting up the two orders is that by starting with speech, it is possible that a linguistic categorical perception can be efficiently applied to gradually degraded stimuli. In the other order, by starting with the "humming," the experimental procedure requires to identify a pragmatic interpretation in a signal lacking recognizable segments and meaning. The increase of linguistic information along the conditions can be informative to observe how perceptual learning takes place and how this affects the categorization process.

The third goal is to address the topic of individual variability in the perception of intonation. For this reason, the experiment is designed to have a good signal to noise ratio of the measures at single subject level and by investigating whether the way pitch is categorized across conditions is related to single subject measures of communicative attitudes, theory of mind competence, and presence of traits belonging to the Autism Spectrum Conditions (ASC).

Overall the experiment aims to rigorously face some of the multiple psychological and linguistic dimensions that are relevant for the categorization of a pitch movement. The data will also be discussed in order to possibly reinterpret previous studies of categorical perception of intonation in different settings, that exhibit a particular variability and plurality of results.

2.2. The categorical perception paradigm applied to linguistic intonation. Are Boundary Tones categorically perceived?

The Categorical Perception (CP) paradigm has been applied to the investigation of pitch modulation at the sentence-level domain mainly addressing two dichotomies: the opposition between emphatic vs. non-emphatic speech and the opposition between affirmative and interrogative sentences, modulated by final Boundary Tones (BTs): prototypically descending in statements, and rising in polar question (Dryer, 2005; Maiden & Robustelli, 2000).

The categorical identification of emphatic vs. non-emphatic speech has been tested in the seminal work of Ladd & Morton (1997). In their hypothesis, emphatic speech was signaled by a larger pitch range variation compared to the one adopted in standard speech (Hirschberg & Ward, 1992) and this could be recognized and identified as a specific pragmatic category. Their results in the identification task indicated the presence of two categories indeed, but the discrimination task was not connotated by the ideal peak in the discrimination function at the presence of the category boundary. A sequence of experiment has led Ladd & Morton (1997;p.339) to formulate an idea that, in my opinion, correctly nailed the whole problem surrounding intonational categories:

> “while we cannot therefore legitimately speak of pitch range distinctions being categorically perceived, it may be useful to think of them as being categorically interpreted.”

The switch from "perception" to "interpretation" entails a core concept: the notion of category has not to be found only at the perceptual level; it lies beyond a debate on whether the signal is only psycho-acoustically perceived or only motorically mapped. On the contrary, there appears to be an integration of sensory-motor information with a high-level cognitive interpretation, that leads to the correct interpretation of the pragmatic message, such as the role of emphasis in highlighting a specific content or referent in an utterance.

Ladd & Morton (1997) formulated this concept considering both the multi-dimensions of the signal and the implication it has, compared to the pure application of the CP paradigm. For them, the categorical interpretation of intonation is built upon a set of continuous acoustic and pragmatic cues that allow the perceivers to rule out one of the two options, in order to decide the category to which the stimuli belong. The factors that can play a role according to Ladd & Morton (1997, p. 339) are "pitch range, voice quality, lexical content, discourse background, relationship between speaker and listener."

Despite the authors’ conclusions, rather ahead of time, in the intonation field the paradigm has been applied and proposed in several languages and slightly different experimental settings, mainly applied to the perception of boundary tones. While emphasis remains a broader concept, with paralinguistic traits, and maybe with a subjective interpretation of the speaker, the application of the paradigm to boundary tones seeks the categories expressing modality (questions vs. statements), which must be univocally interpreted by the speakers of a language.

These studies addressed the opposition between a low boundary tone L%, and a high boundary tone H%, as main cues expressing the pragmatic function distinguishing statements and interrogative sentences (Cummins, Doherty, & Dilley, 2006; Falé & Faria, 2006; Remijsen & van Heuven, 1999; Schneider, 2012; Schneider, Dogil, & Möbius, 2009; Schneider & Lintfert, 2003; van Heuven & Kirsner, 2004).

As a general assumption, the majority of studies tested the binary opposition in languages in which questions are expressed through a H%.

Very few studies reported a "true" CP - presence of a peak in the discrimination function and not of a plateau-like Remijsen & Van Heuven (1999), and (2003), in Dutch. While reporting CP they also reported high variability in the position of the category boundary in the categorization task, but a high consistency in the correlation of individual measures in the categorization with the discrimination rates (r = 0.68).

However, these two studies were completed by a follow-up of Van Heuven and Kirsner (2004) in which the authors tested the existence of three categories - command, continuation, question – instead of the canonical two. The final conclusion to which they came was that the truly categorical opposition was the opposition between "command and non-command"; in terms of signal: "low vs. non-low boundary tone." According to the authors distinguishing between these 2 categories has a real communicative function in speech. This interpretation sustains that an L% signals the termination of a turn and that the interlocutor is not required either to take the turn to add content, as in the case of a question for example nor to wait for further content from the first speaker, as in the case of continuation. On the other hand, according to the authors, the opposition between "continuation and question" is a paralinguistic distinction, since, in both cases, the communicative function of the rise is to signal to the interlocutor to not disengage in the conversation.

Schneider (2003, 2005, 2009, 2012) completed in German a sequence of studies on the categorical nature of the perception of boundary tones (Schneider, 2012; Schneider et al., 2009; Schneider & Lintfert, 2003; Schneider & Möbius, 2005). Along her studies, she shifted from testing purely CP with the CP paradigm to the adoption of the PME perspective, and in final instance she adopted as best perceptual model to describe the perception of intonation, the Exemplar Theory (for a review see e.g., Johnson, 1997; Lacerda, 1995; Pierrehumbert, 2002, 2003).

Schneider & Lintfert (2003) found, in German, a sigmoid categorization response curve combined with a plateau effect in the discrimination function analogous to the effect found by Remijsen and Van Heuven (1999); they interpreted it as CP, since a sigmoid curve could be fitted on the identification tasks' data, but they acknowledged that the discrimination ability within categories was not in line with the original CP paradigm's findings. They hypothesized that a possible explanation for the plateau in the discrimination function could be based on a phonetic analysis of the statement-question continuum. The authors suggested possible individual differences in the position of category boundaries in the categorization task related to the gender of the participants: the turning point occurred earlier along the continuum of stimuli in women's results compared to men's; perceivers that were familiar with the task like phoneticians showed higher turning points compared to naïve perceivers. The authors explained this variability hypothesizing that the specific previous experiences could have modified the perceptual strategies of experienced participants: a more efficient pitch information processing could have shifted the position of the turning point to higher values. Individual variability was also observable in the discrimination task: they mentioned that two of the participants showed to be "pitch listeners" given their higher rate in the discrimination task. In conclusion, the authors supported the possibility of the presence of a third category between statement and question:

“continuation” that might have caused the plateau effect in the discrimination function; they also supported the proposal of Van Heuven & Kirsner (2004) pointing out that the binary opposition could be in function of the factor “terminality”, signaled by a descending boundary tone.

In a following study, Schneider & Moebius (2005) adopted as referring perceptual theory the PME (Kuhl, 1991) and the Signal Detection Theory (Heeger, 1998; Wickens, 2002) and designed an experiment composed by 3 subtests: categorization task, goodness rating, discrimination task. During the categorization task participants were asked to classify the stimuli as either "question" or "statement" or "neither question nor statement." Out of 20 participants, 2 individuals completely did not identify any stimulus as belonging to the third category, and overall the category reached an identification rate of 55%, where the other two reached a minimum of 90%. The authors interpreted this as a sign of disprove of the existence of the third category. Of particular interest is the fact that the identification rate of the category "statement" appeared much clearer-cut and homogeneous than the identification rate of the category question; this was probably due to a too high pitch range of the H%. Besides, a test for the presence of a perceptual magnet was run: by assessing the presence of stimuli rated by the participants as best instances of the categories – the prototypes – and by measuring the discrimination rate surrounding the center of the category. A perceptual magnet was found within the category statement, but not within the category question. In final instance, the presence of the third category – continuation was retained possible, but not fully sustained by the data. A final remark is the fact that the goodness rates for the category "question" appeared to be linked with the presence of high frequencies. Few participants reported that the category "question" sounded a bit unnatural. The lack of a magnet in the "question" category, therefore, could also have been in function of an unnaturally high pitch value of the high tone.

Later, Schneider et al. (2009) performed a similar study but modified the design by adding a context sentence, preceding the target sentence where the manipulation of L% vs. H% is taking place. In their study, there were 3 types of context sentences in function of their right final % tone.

1) L% condition (Context sentence = statement):
   he wants to make a journey. To Panama. / ?
2) H% condition (Context sentence = question):
   he wants to make a journey? To Panama. / ?
3) Wh-L% condition (Context sentence = Wh question ending with an L%):
   What is this? A ticket to Panama. /?

The authors hypothesized that within the identification task, the pitch modulation in the target sentence - To Panama./? - ranging from statement to question, could have been identified in two categories and that the discrimination results would have displayed PME only within the statement category as in the previous experiment. Also, the authors wanted to test if the presence of a question in the preceding context-sentence influenced the position of categories' boundaries in the categorization task, causing the need of a higher F0 to trigger the identification of a "question." No specific prediction was made on the expected differences between the second and third type of context sentences – they both are questions but ending with different boundary tones. Within the categorization task, participants had to choose only between two categories: whether the stimulus was either a statement or a question. In this study, Reaction Times (RT) were also measured.

As overall results, they found that in all 3 conditions the identification task elicited a sigmoid function, meaning the two categories were recognized. The presence of a question in the context slightly shifted the category boundary toward the question interpretation, meaning that a lower rising step in Hertz was required to perceive a question in the target stimulus, after having heard a question. Nonetheless, the shift of the category turning point in function of the context sentence was qualitatively declared, and positioned, along with the 20 steps of stimuli, at about the stimulus 10 when the context sentence was in L% condition, and at about the 11 stimuli. No difference was found between condition 2 and 3. This is rather interesting since the context sentence in condition 3 ends with a final low boundary tone, therefore if the context sentence impacts on the perception of the target, the reason is not be found in the acoustical trace of the final movement that the participants have heard. The RTs matched the interpretation of the sigmoid curve, being higher at the categories' boundaries. The test for the presence of a perceptual magnet revealed a strong effect within the category statement and a smaller one within the category question. Despite the effect in the question category did not appear as strong as the one in the statement categories, the authors supported in contradiction with the previous experiment the presence of a magnet also in the question category.

The two results could also be influenced by an artifact of the CP paradigm. The paradigm can be sensible to the number of proposed categories and the type of task, that can require the identification of one category (e.g., Yes/No paradigm) or the identification of all the presented categories (Kingdom & Prins, 2010). Reducing the available choice from: a) "statement" vs. b) "question" vs. c) "neither question nor statement" to only "statement vs. questions" could have enhanced the presence of a perceptual magnet within the question category.

Finally, in her dissertation, Schneider (2012), summarized her findings. According to the author, the CP paradigm applied to intonation generates, in the discrimination function, a plateau effect instead of a peak aligned to the categories' boundaries. A possible explanation for this difference is identified in the length of the stimuli since stimuli created to test F0 movements perception have to be significantly longer (sentence-level domain) than stimuli created to test CP of segments (syllable-level domain). In the dissertation, the author supported the existence of two prosodic categories "statement" vs. "question" expressed by the two boundary tones L% vs. H%, despite the plurality of results. Moreover, she attributed the variability found in the discrimination rate to the nature of the intonation categories, specifically to their need to be parallelly produced with segmental information carrying syntactic and semantic information. She further supported the idea that the presence of a context sentence, in the shape of a question, influenced the categorization process of the following target sentence and she supported the exemplar theory as the best perceptual model.

In final instance, she proposed that the less categorical perception's effects, found for intonation compared to segments, can be explained by the multidimensional nature of intonation, requiring the evaluation of more than one acoustic cues. The evaluation of these multiple cues and their combination can be possibly done for Schneider at the individual level. She provided the interpretation of segments as "closed" categories – "defined by clear (acoustic) cues that are set early in the language acquisition process" (Schneider 2012, p. 234) and of prosodic categories as "half-closed" categories – their locations, boundaries and centers can be changed or moved in perceptual space, depending on listener's previous speech experiences."

In her view, native language shapes an interval where individual variability occurs, and in line with the Speech Learning Model (Flege, 1995) she postulates that categories can be kept learning throughout the whole life, and the categories' boundaries can be adjusted in function of experience. In this perspective, the position of the categories boundaries is much more subjected to variation compared to what happens at the segmental level.

Moving on to studies that investigated the role of segments in the perception of Boundary Tones, only two tested the presence of CP in function of the manipulation of segmental information. In both of these two studies, the segmental information varied between the identification and the discrimination task, and the identification maintained the clear linguistic information.
Falé & Faria (2006) studied the question statement contrast in Portuguese. They interpreted the controversial results in the discrimination task as mainly dependent from the physical properties of the stimuli: the frequency range (too high H%), the step (in hertz) among the levels, and the duration of each stimulus that is longer in experiments on intonation compared to the experiments using segments. Therefore, they designed an experiment where the identification task was run with segments, in two conditions, one with a male voice and one with a female voice, varying, therefore, the pitch range. The discrimination task, on the other hand, was run cutting the stimuli so that only the final rise, varying in each level was present, and they presented them in a "hummed" version (corresponding to the least linguistic condition in Experiment 1 in this work). The choice was motivated assuming that a shorter duration would facilitate the task and that the "humming transformation" would remove any oddity caused by the cut. The CP experiment was, thus, run by using whole sentences in the categorization task and a hummed boundary tone in the discrimination task. The results of the identification task showed in both conditions a clear sigmoid curve. The results of the discrimination task did not show any peak in correspondence of category boundary; they displayed only a main effect of the order of presentation: the discrimination seemed favored by the presentation of the lower followed by, the higher stimulus (AB order) than vice-versa. This observation is reported in basically every presented study. More interestingly, the overall discrimination rate was lower than 50% and required a step from 2 to 3 semitones. Falé & Faria (2006) discussed the low performance in discrimination as possibly linked to the stimuli they used (the humming) that lacked context. However, they did not specify which specific mechanism of perception is impacted by the presence/absence of segmental information, and they concluded supporting the continuous perception of intonation modulating BTs.

Finally, Cummins et al., (2006) compared the perception of L% vs. H% in 4 conditions from very speech-like stimuli to clearly non-speech (buzz) stimuli, with English material. They aimed to contribute to the debate on the nature of intonational categories by testing the perception differences of speech and non-speech material. According to the authors, the main difference between the speech and non-speech conditions should emerge in the discrimination task; therefore, their interest pertained the trend of the discrimination function primarily. The 4 conditions were composed of one fully linguistic stimulus – the word "Norway"; the 3 non-linguistic conditions were created through the software Praat: a) extracting a single pitch pulse and repeating it over time, and matching it by amplitude with the linguistic condition (voiced set), b) transforming the original signal in a humming stimulus (keeping the 5 formants information), c) transforming the original signal in a "buzz" sound (without the 5 formants information). Their results of the discrimination task reported a significant effect of condition: the less linguistic stimulus (buzz) showed higher correct discrimination rates than other conditions (AX discrimination test) and the presence of a discrimination peak was associated to analogous stimuli-pair in all 4 conditions. The participants performed first the discrimination task in the 4 conditions, then the identification task only in the linguistic condition. The results highlighted the expected sigmoid curve in the identification task and the presence of a peak in the discrimination function that nonetheless did not correspond to the position of the category boundary

in the identification task. The discrimination rate is found higher in the non-speech conditions than in the speech one, but the authors did not furnish an interpretation of the possible cause. They stated that it is unclear whether the discrimination improves because of the absence of segments or because of the spectral properties of the "buzz" stimulus.
The authors supported the hypothesis that discrimination is enhanced simply by the transition from low to high, and not by the presence of a linguistic category boundary, and that therefore the results were in line with the experiments of David House, presented in chapter one, mentioning, in particular, the fact that ramps appeared to be better discriminated than level tones.
This experiment is probably the one that most resemble the experiment 1 that will be presented in this chapter, and it is the first one that directly addresses the interpretation of the curve's turning point, representing or not representing a category boundary, matching or non-matching a peak in the discrimination function. Interestingly it is mentioned that the discrimination peak here does not match the category boundary and it is mentioned that the peak could reflect a psychoacoustic property of the tone, perceived as rising. This can be further interpreted as the observation that participants are not accessing a pragmatic category, conveying sentence modality, while performing the discrimination task; they, on the contrary, evaluate pitch height, and the movement.
I think the experiment shows the interesting fact that a rising tone with and without segments does not show to be discriminated adopting the same strategy, thus the processing could be different; nonetheless, I think that the non-match of the category boundary with discrimination function’s peak could be explainable by the fact that the identification task is run only in the linguistic condition, and after the 4 discrimination sessions. The evaluation of a pragmatic interpretation of the H% and L% might be hard to perform in the discrimination task alone, with 3 out 4 conditions lacking segmental information. If speech and non-speech are not similarly processed, as the literature signals, it might be even harder to match the discrimination peak of non-speech stimuli with the category boundary based on linguistic interpretation performed only after the discrimination task.
Also, the 4 conditions are run in 4 distinct sessions in different days, and this might have introduced noise in the results. It is true that dividing the 4 tasks into 4 independent sessions reduces the effort of the participants and allows for an independent measure, but at the same time, introduces a bias because, in truth, the measures are not independent, if afterward a comparison is observed across categories and not measuring the overall average position of the discrimination peak. It is a repeated measure design, but in this way, it becomes unclear whether the participants were trained by the previous sessions, and whether the results of the first session, were, for instance, poorer than the ones in the last session. Moreover, the number of total participants (6), not balanced for sex, might be not sufficient to clearly model the positioning of the discrimination peak in function of 4 types of stimuli, and it does not really allow to further investigate individual behaviors.

Summarizing, overall the reviewed studies are consistent in showing a stable effect within the categorization task showing that when only two categories are made available to the participants, a sigmoid curve in line with the CP paradigm designed for segments can always be fit. On the other hand, the results do not consistently find a peak in the discrimination function as expected by the paradigm but display either a plateau effect, complete continuous perception with very low accuracy in the discrimination task, or the presence of a peak that does not match the position of the category boundary. The application of the goodness rate, seeking a PME, performed in Schneider's studies seemed to add particularly relevant information. In particular, it appears relevant that: the strong PME effect is found within the statement category, and it is independent of the presence of other contextual information; the instability of the PME within the question category shows to be in function of pitch height and in function of the presence of a question in a previously presented context sentence; the number of categories made available for the participant to choose shows to impact on the results of the identification task, in particular enhancing the formation of a PME within the question category when only 2 categories are presented. The strongest representation of the statement category is found repeatedly in different independent studies, supporting the hypothesis that the pitch normal declination within an utterance is somewhat associated with a non-marked finality interpretation.

Particularly relevant is the suggestion that prosodic categories can be classified as half-closed categories, compared to segments, given the presence of multiple cues co-responsible for a unique pragmatic and syntactic interpretation. The multidimensionality of the signal seems to play a more relevant role as the pitch range increases. For anatomical reasons the discrimination sensitivity decreases as the frequency of pitch increases (see the concept of Critical Bandwidth in chapter 1), therefore the linguistic organization carrying redundant cues at different levels (e.g., syntactical, pragmatical) helps disambiguate could function as support in cases where the perception of a single cue alone is hardly clearly identifiable as belonging to one specific category.

2.3. Categorizing and The Multistore Model of Memory

Recently, Holt & Lotto (2010) and Lotto & Holt (2016) and Heald, Van Hedger, & Nusbaum (2017) addressed the debate on the categorical perception of speech, focusing mainly on segments, pointing out a key distinction: investigating the presence of categorical perception alone is not sufficient to explain the categorization process. In line with the logical considerations on the abstract notion of category provided in Chapter 1, paragraph 1.3, Holt & Lotto stress that the first task in a CP paradigm often described through the interchangeable use of the words: "categorization", "identification", "labeling", is in truth only a task of "identification" and not of "categorization" if the categories are perceptually already available, and the incoming stimulus is compared to a similar already existing entity. On the contrary, the word "categorization" specifically refers to the generalization process occurring during the perception of several entities. Adopting the line of reasoning in Chapter 1, the distinction introduced by Holt & Lotto allows to disentangle the processing of "Object-sameness", whether an incoming stimulus belongs to a category or not, to the processing of "Relation-sameness"(Robert, 2005), whether several incoming stimuli are groupable in a category that differs from a different one for a set of distinctive features.

This is formalized explicitly for auditory perception as the concept of

> “identification” meaning “a decision about an object’s unique identity that requires discrimination between similar objects” (Holt & Lotto, 2010; p. 5 – Authors’ manuscript) opposed to “categorization” meaning “a decision about an object’s type or kind, requiring generalization across the perceptually discriminable physical variability of a class of objects” (Holt & Lotto, 2010, p. 5 & 6 – Authors’ manuscript)

By shifting the perspective from seeking the presence of CP to understanding how categorization takes place it is possible to account for perceptual learning, and it is possible to account for the possibility of categories to form, thus, to investigate in the field of intonation the shift from the auditory perception of a rising tone to the formation and recognition of the linguistic, pragmatic interpretation of a rising tone – e.g., the question.

This process grants a much larger modelling flexibility than for example the classical approach described in Chapter 1, because it allows first to overcome a strict division from speech to non-speech perception, pitch

with and without a linguistic function can be successfully categorized, and there is nowadays a broad evidence that CP is not a matter of only speech (e.g.). Secondly, it allows to account for the formation of a new category in a second language; thus it allows to study the evolution of its formation in the interlanguage. Thirdly, it can account for the different stages through which the representation of a forming category is shaped and reinforced, thus for the occurring of the comparison of several trials through extrinsic reference that is necessary when more than one parameter is present in the stimuli, and the categorization process is in function of a parameter changing in time within a single stimulus, as it can be for pitch movements aligned on vowels. Finally, it allows fitting the multidimensional nature of intonation, requiring the check the presence of several features such as the alignment of specific contour shapes to specific metrical positions, on a meaningful portion of discourse. At the initial stages of a category formation, it is possible that the categorization is carried on considering a partial amount of information, the features considered more relevant or simply perceptually more salient, and that later on the properties of more levels are integrated, improving the representation of the category.
Within this perspective, performing the identification task within a CP paradigm implies:

1. The activation of existing categories which will constitute a template (Top-down activation)
2. The evaluation of the incoming stimulus in function of the template (Object-sameness)
3. In cases of failure of the identification process, the resetting of the evaluation of several incoming stimuli through extrinsic reference (Bottom-up activation)
4. The organization of several incoming stimuli in an emerging category (Categorization process)
5. The positioning of the new category in within a perceptual space comprehending the already existing ones (Relation-sameness)

In a cognitive perspective the two main mechanisms responsible for the computation of these operations are mechanisms of selective attention (Francis & Nusbaum, 2002) seeking the relevant features of the existing categories in the incoming stimulus (Object-sameness), and a mechanism able to keep available for comparison the immediately perceived stimuli, allowing a bottom-up generalization process.
Since in the experiment in this chapter focuses on the role of segmental information in the evaluation of a rising Boundary Tone (BT) for the ultimate recognition of sentence modality, investigating how modality emerges starting from the absence of sufficient segmental information, I present here a memory model developed by Xu et al. (2006) to explain the category formation of a rising tone (ramp) in speech and non-speech conditions, perceived by native speakers of the specific tonal language, and native speakers of a different intonational language. The model, thus, is designed to account for the perception of pitch with a lexical linguistic function by speakers of a language in which that function is not present, therefore for the cases in which the incoming stimuli do not match any existing representation.
The Multistore Model (MM) has been proposed in order to explain CP and quasi-CP results, depending on language experience. In their experiment, Xu et al. (2006) compared the categorization process of level tones (flat contours) and ramps (rising contours) in native speakers of Chinese and English. The tones were aligned to meaningful speech material in Chinese in one condition, opposed to a non-speech condition. Since the distinction of level tones and ramps is a distinctive feature only for the native speakers of Chinese, the experiment could account for the influence of native language in speech and non-speech processing of the same pitch contour. Nonetheless, differently from what will be presented here, in this experiment the task required specifically to judge the acoustic dimension of the contour, not the linked semantic representation.

The results of Xu et al. (2006) showed the presence of a peak corresponding to the categories' boundaries in both conditions (speech and non-speech) in the group of native speakers of Chinese, and the presence of higher discrimination scores in the non-speech condition compared to the perception of linguistic stimuli in the results of both the groups of speakers. Within category, native speakers kept a higher discrimination rate compared to Chinese native speakers in both stimulus types, and in both language groups, discrimination rates were higher in the non-speech condition compared to speech.
The authors proposed to explain these results with a language-specific attention related mechanism (the recognition of Chinese tones) impacting on both linguistic and non-linguistic perception. In a very broad view, Xu and colleagues proposed that CP effects arise whenever participants are required to judge stimuli sharing features similar to linguistically relevant ones in their referring linguistic system, independently from the contextual presence of speech or not and formulated the MM model.
The model assumes that CP stems from at least two memory processes as described by the Dual-Process Model (Fujisaki & Kawashima, 1969, 1970, 1971; Pisoni, 1975) and the Signal Detection Model (SDT)(Green & Swets, 1966; Macmillan & Creelman, 1991), but in addition to the two previously mentioned models, the MM considers "categorization a process inherent to perception" (Xu et al. 2006, p.1070). In their view, auditive categorization can happen already at early stages, and it is a process independent from speech. The idea is that categorization process, and thus, the early perceptual internal response obeys discrete probability distributions modeled by a distribution resembling a sigmoid function. This response for the authors is the mechanism that in the classical view Liberman and colleagues attributed only to speech, and the response enhancing the discrimination on the basis of "acquired distinctiveness." Given the fact that CP has been found in several domains outside of speech (e.g. perception of music intervals (Krumhansl, 1991), non-speech continua (Miller et al., 1976), human faces (Beale & Keil, 1995), perception in animal cognition (Kluender, Lotto, Holt, Greenberg, & Ainsworth, 2005; P. K. Kuhl & Miller, 1975, 1978) Xu et al., try to extend this mechanism to a general sensory-perception modality, valid for instance also in the visual domain.
According to the authors, the MM represents an advancement compared to the Dual-Process Model and compared to the SDT, given the fact that the former model is considered limited as applicable only to speech material, and the latter one is found not to able to distinguish the storing of stable representations, existing prior to the exposure time, and that are stored in long-term memory (e.g. native language), and the storing of the several stimuli, presented during the exposure time, that are in course of categorization (e.g. the total amount of stimuli presented in an experimental paradigm). This storing distinction is fundamental for them since it is the mechanism allowing for perceptual learning starting from the first exposure to any kind of stimuli.
Within the MM model the input is firstly processed in a sensory memory trace, containing non-analyzed "raw" data, retaining information for a span lasting about 300 ms, and then it can parallelly be processed in two storage devices. The "raw" stimulus can reach an Analyzed Sensory Memory (ASM), which can retain information for a span lasting in the order of seconds. The ASM is the information storage allowing for extrinsic reference, and the evaluation of contextual information, and contains fine-grained analyzed information about the phonetic details of the contour. Parallelly, the features found relevant for a categorization process are retained in a Short-Term Categorical Memory (STCM). The STCM can communicate with the long-term memory store containing the stable and permanent discrete representations of perceptual categories, and categories formed during the time of exposure can be preserved there. Permanently stored categories can be matched by the bottom-up incoming stimulus but also can guide selective attention in a top-down activation of a template.
In the discussion of their experiment, Xu et al. (2006) explained the effect of native language on perception as a top-down activation of the existing categories "level tone" and "ramps," present in Chinese, and not

present in English. According to the authors, English speakers, not having an available template, instantiated a perceptual learning process, making use of the STCM. This mechanism for Xu and colleagues is the mechanism generating quasi-categorical perception, instead of the classical CP. The fact the raw stimuli are parallelly "categorically" and "continuously" processed generates the discrimination advantage in the presence of a boundary, but both the fact that the category is in the course of formation and that continuous perception details parallelly come from the ASM, the discrimination within category remains possible, and this results in the presence of the plateau effect.

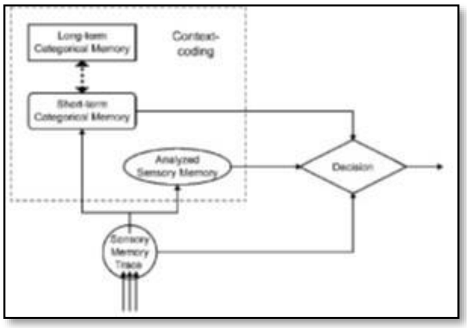


Figure 2.3.1 Multistore Model - adapted from Xu et al. (2006)

The MM elegantly describes and accounts for the previously presented steps occurring during the categorization process. It accounts for both the presence of a native language phonological and phonetic representation, for a linguistic and a non-linguistic decoding of pitch, for the formation and reinforcement of a new category in the course of the experiment, through bottom-up evaluation, and it can adequately explain the processing of extrinsic reference.

I find of particular interest the double parallel mechanism allowing for both the processing of continuous phonetic features and for an early stage quick categorization storage that overall has a retaining time of the order of seconds. The processing of intonation as described in Chapter 1, requires the integration of several levels of contextual information and the evaluation of slow frequency transitions. The research adopting the Electroencephalography (EEG) technique in particular, has highlighted an expanded span of time reaching almost 2 seconds for the integration of a sentence-final boundary tone with the processed informational content and the evaluation of the pause signaling the phrasing of the prosodic unit. The specific Event-Related Potential (ERP) signaling this type of processing is called Close Positive Shift (CPS) (see, e.g. Steinhauer, 2003; Steinhauer & Friederici, 2001). It seems therefore plausible that it exists a mechanism able to retain a set of cues considered relevant for a final evaluation for the amount of time of the order of seconds. Moreover, it interesting to assume that while an early categorical representation becomes available, analyzed data based on continuous perception also remain available to the system. Adopting a model like the MM could help to resolve the existing tension in the literature between two main approaches. In fact, the model could be an answer, on one side, to the approaches that see in the abstract discrete entities the presence of a linguistic system that comes in contact with sensory information, and presumably filters the perception (see: the early account of the Motor Theory of Speech Perception, and see the strong theoretical division that the Autosegmental Theory of Intonation makes between phonological representations and phonetic implementation); on the other, it can account for the evidence actually coming from experiments in perception that strongly signal a determinant role of alignment (a phonetic dimension) in the identification of intonational categories. Also, adopting a model like MM could allow fitting the perceptual evaluation of L2 categories, that the presented L2 acquisition models (Chapter 1, paragraph 1.5.2.) also reconduct to the phonetic level.

The MM model is adopted as baseline mechanism in this experiment to generate specific predictions on the categorization process of stimuli presented in the order that goes from low recognizability of segments, to fully recognizable and conveying meaning segmental information. I hypothesize that a mechanism sufficiently resembling the processing described by the MM can be at the basis of the categorization process of a pitch contour, lacking relevant contextual information.

Therefore, the following experiment is designed considering both the evidence coming from the experiments adopting the CP paradigm applied to intonation, introduced at the beginning of the chapter, and the processing model granted by the MM. The literature addressing only the presence of CP will be a reference for the interpretation of the results in terms of presence or absence of the two categories expressing sentence modality (statement vs. question); the MM model will be a reference for accounting for the formation of a new linguistic or non-linguistic category associated with a rising final pitch movement.

2.4. Experiment 1

On the basis of the previously presented literature, the following study is composed only by the categorization task of the classical CP paradigm, with the aim to investigate the role of "pitch co-occurring linguistic and segmental information", that it is labelled here as "co-text". The task requires in every condition to categorize pitch movements as boundary tones signaling the modality change: statement vs. question by means of a binary forced (2AFC) linguistic decision.
To capture the influence of the omni-variable "co-text" on the pitch categorization process the variable can assume 4 different levels:

1) Full recognizable and accessible linguistic information (word condition)
2) Partially recognizable linguistic information – deprivation of semantic access (pseudoword condition)
3) Partially recognizable and partially accessible linguistic information – deprivation of semantic access and of phonematic full recognition (foreign condition)
4) Unrecognizable and inaccessible linguistic information – full deprivation of linguistic information (voice information still available – recognition of vocal tract acoustic) (humming condition)

The 4 conditions are ordered in two directions generating two experimental sequences: *Sequence 1,* in which co-text is modulated from accessible linguistic information to inaccessible linguistic information, and *Sequence 2* in which co-text modulates in the reversed order, and participants are exposed first to stimuli where language recognition is experimentally impaired. At the end of each sequence the first condition of the sequence is represented as Retest phase. Thus, respectively, within Sequence 1 in 5th experimental position the fully linguistic stimulus is represented to the participants, and within Sequence 2, in 5th experimental position the less linguistic stimulus is represented to the participants.
Given the previously described role of language, I hypothesize that the categorization process of the fully linguistic stimuli will not be affected by the other conditions available in sequence, and this will be seen by comparing the values of the parameters determining the sigmoid function generated in the 1st and 5th condition in Sequence 1. On the contrary, the categorization process of a F0 modulation deprived of co-textual linguistic information is hypothesized to vary in function of the co-text levels within Sequence 2, and therefore the comparison of the sigmoid curves' parameters at the beginning and end of the procedure are expected to display a significant difference.
The choice of investigation of BTs is motivated by the fact that their opposition allows to clearly disentangle statements from questions is several varieties of Italian, therefore it is conceivable that the two categories are clearly available to the participants. Despite some of the cited studies (e.g. Schneider & Möbius, 2005; van Heuven & Kirsner, 2004) tested the possibility of a third category occupying the perceptual space between L% and H%, the contrast is kept with the two original levels. The adoption of a 2AFC paradigm is motivated by the fact that the interest of this study is not in looking for the presence of categorical perception but on the psychophysical modulation of the sigmoid logistic function depending on the degree of available segmental linguistic information in the stimuli. Adding a third category would have further complicated the experimental design. In second instance, much controversy is still present in studies with two categories only, and given the explorative nature of this experiment, it has been considered advisable to not further complicate the design. Finally, the decision to allow participants to choose between two given labels (statement and question) and not as suggested by the study of Van Heuven and Kirsner to let participants choose between "statement and non-statement", is motivated by the fact that it would have

generated a yes/no design centered on the statement category. Independently from the linguistic theoretical discussion, a yes/no task can bias results by means of a conscious and unconscious strategy of the participants (Kingdom & Prins, 2010) to favor “yes” answers (in this case: statements), or paradoxically can induce the opposite effect to push participants to adopt a stricter criterion in order to properly answer “yes”, shifting in this case to the left the category boundary and being more restrictive towards stimuli close to the category boundary. 2AFC designs can also be affected by biases, but the two alternatives are proposed “with the same weight”. In final instance, within communication, the concept of (polar) question clearly exists in the mind of every speaker; in cases of uncertainty, in everyday dialogues, speakers commonly ask: “was it question?” or: “were you asking me?” and, rarely, in normal conversations a speaker would disambiguate asking “was It not a statement, was it?”. Given the fact that in this design the linguistic information gradually degrades, the linguistic function associated to the categories is voluntarily left completely and clearly available to participants. The F0 modulations are aligned to “word-long” and not “sentence-long” segmental material in order to diminish the impact on working memory. Therefore, in the experiment participants perceive holophrasis.

Overall the experimental setting is composed by a between-design of the 2-levels factor “order” generating Sequence 1 and Sequence 2; in each sequence the co-text factor generates a 4 levels within-design; in adjunction the condition in position 1 is retested at the end of each sequence, generating a test-retest condition analyzed independently.

Sequence 1			
Order of presentation	Condition	Code	Stimulus
1	Existing Word in Italian	word - test	/a’nemoni/
2	Pseudoword in Italian	pseudoword	/e’nimena/
3	Pseudoword in Russian (with Italian syllabic structure)	foreign	/l’linieli/
4	Vocal stimulus - Humming	humming	
5	Existing Word in Italian	word - retest	/a’nemoni/

Table 2.4.1 Order of the experimental conditions in Sequence 1

Sequence 2			
Order of Presentation	Condition	Code	Stimulus
1	Vocal stimulus - Humming	humming - test	
2	Pseudoword in Russian (with Italian syllabic structure)	foreign	/l’linieli/
3	Pseudoword in Italian	pseudoword	/e’nimena/
4	Existing Word in Italian	word	/a’nemoni/
5	Vocal stimulus - Humming	humming - retest	

Table 2.4.2 Order of the experimental conditions in Sequence 2

2.4.1. Logistic function and parameters of interest

Data point obtained by the response of each participant in the categorization task in each condition are fitted to a logistic function given as:

$$F_L(x;\ \alpha,\beta) = \frac{1}{1+\exp(\beta(\alpha - x)}$$

Within this work the β parameter is named "slope" and the α parameter is named "midpoint".
The slope value can range from a minimum of 0 to the limit of ∞, in practice it never assumes values greater than 1, the midpoint value corresponding to frequency steps manipulated in the boundary tone can span overall between 0 and 140 Hz and corresponds to the half of the range at 70 Hz.
The fitting procedure is obtained by using the Palamedes Toolbox on MATLAB platform. The other two parameters, required by the fitting procedure, γ (guessing rate) and λ are set as constants equal to 0, because it is possible that the procedure never forces the participant to find a right answer in presence of two possible choices in complete absence of a criterion. The two linguistic categories, statement and question, are assumed to exist for every perceiver native-speaker of the selected variety, and the uncertainty generated by ambiguous pitch contour is the core of the experiment. It is therefore also assumed that the perception of prototypical statement-contours and prototypical question contours can reach the asymptotic values of 0 and 1, corresponding to 20 out of 20 participant's responses.

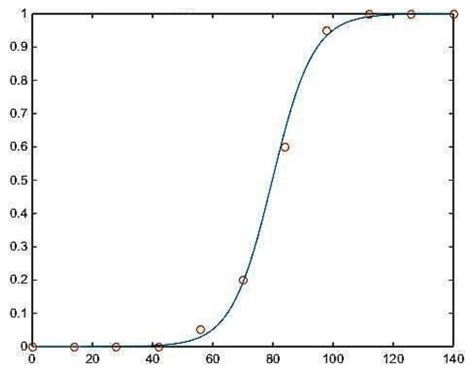


Figure 2.4.1Prototypical fitting of the logistic curve on a participants' responses

Throughout the whole experiment the curve will always be depicted in function of the responses to the category "question", therefore signaling on the bottom left of the figure the percentage of 0% of "question" judgements, and on the top right the 100% responses. Whenever the midpoint shifts towards higher values on the right side of the plot, it will indicate that a higher range of frequencies is needed to change the judgement from statement to question, and vice versa.

2.4.2. Hypothesis

In sequence 1, starting with condition 1: the full recognition of segmental information and lexical content must be reflected in slope values and the positioning of the midpoint. In this condition, the midpoint must reflect the category-boundaries set on the basis of the native language system; the slope parameter must reflect the degree of certainty with which the decision is made: a steeper slope would reflect a very sharp distinction between the two categories that are recognized with high certainty. Moreover, it would highlight that the range of frequencies is sufficient to reach the correct identification and allows the category shift. A shallow slope, on the contrary would signal high indecision in the placement of a boundary, and the need for a broader range of values to undoubtedly switch the category. In absence of the discrimination task, perception is considered more categorical in presence of steeper slopes, since it implies the presence of a clearly identifiable category turning point (midpoint).

Along the conditions, the categorization task should become harder since the segmental information decreases. I expect therefore that in condition 2 (pseudo-word) and condition 3 (foreign-word) slope values will decrease, reflecting the role of segmental information in the decision-making process, raising uncertainty about the presented stimuli. In condition 2 and 3, midpoint values should increase reflecting uncertainty and causing the need of more trials to correctly determine the categories boundaries.

In condition 4 (humming), where no segmental information is available but the same linguistic task of identifying sentences and questions is required, the hypotheses are twofold: if the condition is processed in the same manner as previous presented stimuli, then the midpoint value should linearly increase in line with condition 2 and 3, and the slope values should decrease. On the other hand, the literature highlights in tasks requiring to auditorily identify the contour, and not the pragmatic function, that pitch processing without segmental information is found to be easier compared to perception when segments are present (House, 1990; Xu, Gandour, & Francis, 2006). In Xu et al. 2006, the lower short-term memory load required to process non-speech material, compared to speech material, is observable in lower midpoints in the non-speech conditions.
In this study, therefore, if the absence of segmental information simplifies the task and changes the way in which pitch is processed, I expect the midpoint of condition 4 to be lower than midpoints in condition 3 and 4, and if condition 4 completely fails to trigger the linguistic processing, then it is also expectable a value lower in this condition than in the word condition. Analogously, the change in processing and absence of segmental information should also emerge in the modulation of slope values, that reflecting an easier task, should increase compared to condition 3 and 4, and possibly also compared to condition 1.
In Sequence 1, condition 2, compared to condition 1, increases the cost of identifying a pragmatic function without a link to a meaning. The shift from condition 2 to condition 3 reflects only the cost of decoding of segmental information. From condition 1 to condition 3, which represents a highly possible ecological case when I perceive speech in a language I don't master, differences in slope and midpoint values reflect then the sum of higher processing costs in segmentation and absence of lexical access; a change in slope and midpoint parameters from condition 1 to 3, will highlight then that the processing of pragmatic function of pitch requires both phonological and lexical recognition.

In Sequence 2, the variable co-text will vary from unrecognizable segmental information to fully recognizable segmental information (the task requiring to identify the categories of statement vs. question will be kept the same). In this sequence, if segmental and semantic access play no-role in the classification process of pitch then no difference should be found compared to slope and midpoint parameters found in sequence 1; on the contrary I argue that segmental and lexical information is required from top-down activation and that the missing information generated by this experimental order of conditions will cause a strategy-change and the instantiation of an extrinsic reference process of the incoming series of stimuli.

I expect the pitch movements in condition 4 (humming) now first in order, to be hardly categorizable on the basis of a linguistic criterion, but easily categorizable on the basis of acoustic features only. If a pure linguistic strategy is set and not changed by participants, then slope values should be the lowest and midpoint values the highest, reflecting uncertainty in the decision process, and the need of multiple trials to establish a possible anchor and category boundary. If participants can convert the pragmatic task in an acoustical one, basically associating the question category with the highest possible F0 values in the stimuli and therefore internally change the task to an acoustic low vs. high pitch distinction, then the slope must be higher than in any other condition, and the midpoint can be either set in the exact half of range, or, as Xu et al. 2006 suggests a criterion for easiness in task-computation, it can be the lowest possible midpoint value of the whole sequence.
Condition 4 in first place should then set an anchor point for the developing of a bottom-up strategy along the conditions reflecting in this case the categorization process defined by Lotto & Holt (2010, 2016), and not an identification process generated only by the presence of mental representations of the categories question and statements, expressed by H% and L% BTs, top-down activated.

Condition 3 (foreign-word), being presented in second place, represents now the first condition having segmental information, but nonetheless, it will not be fully recognizable. The change from condition 4 to condition 3 can help to shed light on the nature of the processing that occurred in the previous condition.
If the process started already as linguistic in condition 4, then in condition 3, slope should linearly increase, and the midpoint should linearly decrease, since the criterion necessary to compute the task is clearer, fewer trials are needed to be certain about categories-boundaries and the midpoint should start increasingly shifting toward the position of the true linguistic category boundaries.
On the contrary, if in condition 4 the decision is made only on the basis of the perception of acoustic low vs. high pitch contour, then the introduction in condition 3 of segmental information should change the strategy, therefore in condition 3 the slope should abruptly decrease signaling the cost of the task, and the midpoint should be the highest among the conditions where the segmental information is available.

Condition 2, following condition 3, should cause parameters to change linearly: the slope values will increase, the midpoint values will decrease.

Condition 1 should then reflect the fully linguistic activation and it will represent the matching of bottom-up generalization processes with the top-down activation of question and statement mental representations. Therefore, in sequence 2, condition 1 (word) should have the highest values of slopes among the linguistic conditions, and the lowest values of midpoint.

The retest condition of “word” in sequence 1, and of “humming” in sequence 2 is a control condition to observe whether participants maintained a sufficient level of attention along the experiment, and to verify whether the sequence has functioned as a perceptual training.
Correlations should be found between the condition 1 at the beginning of sequence 1 (word-test) and the condition 1 at the end of the sequence 1 (word-retest).

I expect the slope and midpoint values to correlate between word-test and word-retest conditions and not to significantly vary at the beginning and end of the sequence. The variation is not expected because the degree of certainty about the judgement reflected in slope values, and the category boundaries reflected in the midpoint values, must be set by the top-down strategy fully activated by the word condition both in test and retest phase. On the other hand, in sequence 2, the humming-retest condition appears after the word condition at the end of a categorization process last through the whole sequence. In this case I expect perceptual learning's effects, and I expect that the humming-retest condition is affected by the full linguistic activation of the preceding word condition.

If, in sequence 2, the humming-test condition were processed on the basis of bottom-up psychoacoustic criteria, in the humming-retest condition, the values of slope and midpoint should resemble the values present in the humming condition in sequence 1 (condition 4), or reflect an even refined strategy, therefore the slope in humming-retest should be steeper than in the humming-test condition. The midpoint values should correlate with the preceding word condition, transferring on the humming-retest condition the linguistic category boundaries. If, in sequence 2, the humming-test condition were processed adopting a linguistic criterion, and this should be observable looking at the change between the humming-test condition and the foreign condition, then in this case in the humming-retest condition the perceptual learning process together with the added linguistic segmental information, should let the slope values linearly increase and the midpoint values linearly decrease.

2.5. Method

2.5.1. Stimuli Elicitation and Recording

8 speakers (4 females 4 males) from the area ranging from Trento to Padova in Northern Italy were recorded at the University of Padua in a sound-proof room, through a Shure SM 58 microphone at a sample rate of 44.1 kHz. 1 female native speaker of Russian was recorded at the University of Trento in a silent room at a sample rate of 44.1 KHz.

2.5.2. Real-words and Pseudo-words elicitation

In order to elicit natural stimuli, I developed a guided procedure requiring to utter single-word utterances as semi-spontaneous speech. The procedure was developed to elicit real words and pseudo words. The procedure consisted of slides where the speaker could read either a "task" or a question he/she had to reply. A visual stimulus representing the task, or the answer was presented in background, in order to avoid a strict, read-speech effect, or on the other hand complete free speech. Participants could read the target utterance underneath the image. Speakers were instructed through some practice trials to mentally read the task, think about the answer while looking at the picture, and utter the target sentence,
In the following picture:
Task: "Look at the picture and tell me what you see"
Guided Answer: "Anemoni". “Enimena”.

Figure 2.5.1 Example of the material used during the elicitation phase

The guided procedure was designed to elicit several sentences, in different contexts and of different focus types. It entailed also different lexical choices as target referents in order to avoid a strict repetitive task. Out of the corpus, the target real-word utterance /a'nemoni/, and the target pseudo-word utterance /e’nimena/ were selected as baseline stimuli. Both stimuli display the homologues phonotactic structure, composed of 4 syllables with word-stress on the antepenultimate, sufficiently distant from the boundary tones. Both selected stimuli were clearly uttered from a female speaker as statements.

2.5.3. Foreign Russian pseudo-word elicitation

I selected as a referring foreign language, Russian, because it is a language rarely studied within mandatory education in the Italian system, therefore, it is a language to which the majority of the Italian population is rarely exposed to. A combination of consonants and vowels not available in the Italian phonological inventory, but still recognizable as speech, was selected to construct a pseudoword sounding foreign.
The selected phonemes were precisely:

- A close central unrounded vowel: / ɨ /
- A mid central unrounded vowel: /ɘ/
- An alveolo-palatal nasal consonant: /n̠ʲ /
- A dental Alveolar Lateral Approximant: /l/
- The phonotactic properties of the target word /a'nemoni/ were kept constant constructing the foreign pseudo-word: / ɨ ' l ɨ n̠ʲ ɘ l ɨ / (l'linieli)

The stimulus was recorded by a female Russian native-speaker (L2 speaker of Italian, aware of Italian phonotactic structure) only in statement condition after hearings several times the target word /a'nemoni/ in Italian. Since the word is non-existing in Russian and the stress pattern differs from the Italian one, several repetitions have been recorded and durations have been readjusted in the elaboration phase.

2.5.4. Stimuli manipulation

The target stimulus was selected among the several recording of the 8 participants, meeting clarity of speech criteria and a speech style natural and usual in the selected northern regional area of Italy. The average length of the single words "anemoni", and "enimena" was measured, while they were uttered by speakers within sentences and as single words, in affirmative sentences and information seeking questions. Both affirmative and interrogative words-length was considered in order to adopt a stimulus duration that could fit both descending and rising pitch contours without creating significative distortions. The recorded words entries had an average length of 525 ms and the 4 syllables averagely respected a duration proportion of respectively: 17% 31% 26% 26%.
The starting stimulus for the manipulation was the pseudo-word "enimena". The original audio file was 523 ms long. The syllable duration was readjusted on the basis of the average proportion previously calculated, reaching a syllabic duration of: 89 ms - 162 ms - 136 ms - 136 ms. 50 ms of "silence" were added before and after the signal. All stimuli have been created following this prototype, and pitch has been manipulated in all conditions on the very same phonotactic structure. All stimuli equalized at 70 dB scale intensity.
Before creating an artificial boundary rising the base pitch profile was adjusted and cleaned from minor irregularities. Pitch points outside the syllable (voicing range) were deleted. The original F0 peak present at the beginning of the utterance aligning from the pretonic to the tonic syllable was lowered in range, creating a less steep descending contour in the nuclear pitch accent H+L*. This choice was adopted in order to avoid the Gussenhoven-Rietveld effect in combination with the final artificial rising, that would have had created a marked unnatural question contour. The final boundary tone L% was smoothed in order to sound natural but to be “flat” in the most possible way.

The prototypical statement contour assumed therefore the following shape: H+L* L-L% ranging from 253 Hz to 200 Hz.

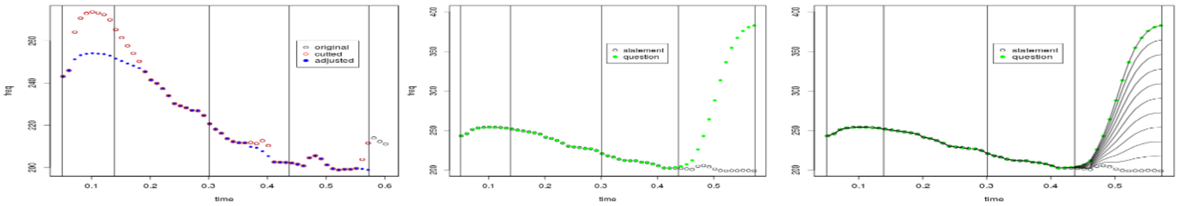


Figure 2.5.2 Schematic representation of the generation of the 1000 artificial rising contours

The artificial final rising boundary tone was created by adding a parabolic function (tapered with a hyperbolic tangent with half rising 60ms after the onset of the last syllable) with an offset from the original pitch value of 140 Hz. This value, as well of the shape and steepness of the rising, was qualitatively drawn based on the observation of the rising in the question production of the same speaker and to undoubtedly convey the question interpretation, a peak height about 1,3 times higher than the peak present at the beginning of the utterance was created.
The prototypical question contour assumed therefore the following shape: H+L* L-H% ranging from 253 Hz and 200 Hz in Pitch Accent position and from 200 Hz to 340 Hz in Boundary Tone position.
The last step consisted in the exportation of 1000 stepwise levels of the rising, and it was done linearly. Each step roughly corresponded to an increase of 0.14 Hz of the amplitude of the final rising. Here Only 10 steps are drawn for visualization purposes. 1000 files ranging from prototypical affirmative intonational contour to prototypical information seeking polar question contour were created.
The 1000 Pitch contours were applied to the 4 segmental conditions through the software PSOLA in Praat (Boersma, 2001): to stimuli “anemoni” and “enimena” uttered by the recorded Italian speaker, to the Russian stimulus, and to a sound generated in Praat – referred as “humming sound” – not displaying any recognizable segmental information.

The transcribed resulting 4 conditions consisted then in :

- Real Word: *anemoni*
- Pseudo Italian Word: *enimena*
- Pseudo Russian word: *ilinileli*
- Humming

The final prototypical pitch contours for the statement category and for the question category were the following:

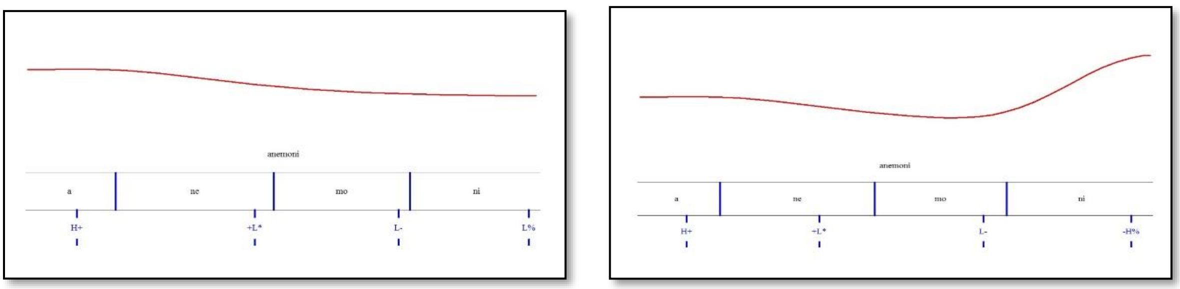


Figure 2.5.3 Schematic representation of the alignment of the two categories' prototypical contours

2.5.5. Stimuli presentation

In each sequence, the experimental conditions were presented in the previously described fixed order. In each condition 11 levels were created from the 1000 stimuli battery. From level 1 belonging to clearly "statement" having a L% with F0 = 199 Hz, each level was selected every 100 artificially created steps, corresponding to 14 Hz. Each level was presented 20 times generating a total amount of 220 stimuli. Stimuli within condition were presented in a random order and were divided in 5 blocks of 44 stimuli each.

2.5.6. Participants

All participants were informed prior the experimental session about the overall procedure and about their right of leaving at any time the experimental session, if uncomfortable, with no repercussion. All participants signed a written consent of voluntary participation before participating to the study. No form of monetary reimbursement was given to any participant, in the case they were undergrad students of the faculty of Cognitive Sciences, the overall participation's period of time was counted and recognized as experimental session time and could therefore be summed up and converted in useful credits in their study-plan.

37 participants took part at the experiment. The participants are all women, native speakers of Italian, of age between 19 and 27 years old. One speaker is 48. Participants have all been raised as monolingual speakers, in northern-east of Italy. 19 perceivers of 37 coming from Veneto, 1 from north-east Lombardia, 1 from Emilia Romagna, 8 from the province of Trento (Italian speaking province of Region of Trentino-Alto Adige in Northern Italy). No one has reported hearing deficits.

Participants had to fill in prior to be physically called in the lab, an online screening questionnaire. Participants were required not to have any particular musical education nor a specifically high linguistic education, resulting in advanced proficiency levels in more than one language. In order to form a sample of population representing individuals without any particular relevant training for F0 discrimination abilities. Since English as a second language is mandatory in Italian education system, participants speaking English as L2 were accepted.

Proficiency levels were self-declared following the Common European Framework of Reference for Languages (CEFR) in 5 language-aspects: listening, reading, interaction, oral production, and written production competences. For each aspect participants, with the referring table explaining the actual correspondence between CEFR level and language skills, were asked to state their level ranging from the labels: "without any competence", A1, A2, B1, B2, C1, C2 language level. The total amount of participants responses has been added, and the average number of participants stating the same competence level in each aspect of each language has been calculated. The total number of answers has been averaged per language and converted in percentage for comparison.

Out of 37, one participant declared C1 and C2 levels in Spanish and French and reported she had been living abroad. No participant reported to have concluded the higher musical education cycle, therefore no participant exhibited a recognized certification for musical competences; two participants self-declared a private training in a musical instrument or singing for a period of time of 10 years or more. These 3 participants have been excluded from the sample because not meeting experimental criteria.

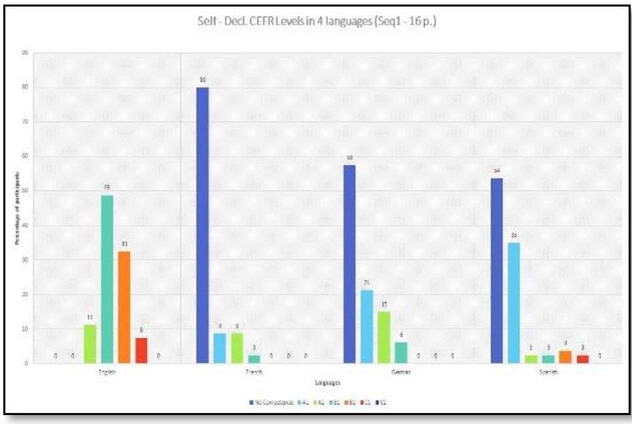
The experimental sample resulted in 34 participants.

For what concerns the use of local dialect:

6 participants (18%) reported a minimal linguistic competence in their regional dialect in comprehension, and no use at all in production.

12 participants (35%) reported to not actively speak their regional dialect but to be extensively able to understand it.
12 participants (35%) reported to use dialect in a productive manner only in a friendly and well-known environment, being able to produce simple sentences.
4 participants (12%) declared a broad use of dialect, declaring a comparable linguistic competence to Italian. one of the four is the 48 years old participant.
The 34 participants were randomly assigned to the two experimental sequences.
16 belonged to the experimental group and phase called: Sequence 1; and 18 belonged to experimental group and phase called: sequence 2.

Sequence 1 **Sequence 2**



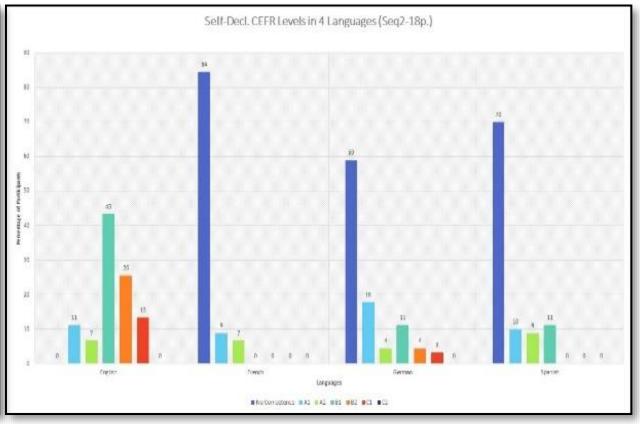


Figure 2.5.4 Distributions of L2s spoken by participants in Sequence 1 and 2

2.5.7. Procedure

The experiment was run in laboratory of the Department of Cognitive Science, of the Faculty of Trento, in the town of Rovereto in Northern Italy. Participants were seated in individual sessions in front of a computer in a quiet room at the presence of a research assistant. Each participant listened to the audio files through headphones (Sennheiser hd230) and with the support of an external Sound Card (Echo Audio Fire 2 - Firewire 400). The procedure was launched on the MATLAB platform (2014b), with the aid of two specific toolbox: Psychtoolbox (3.0.12.426025497) and Palamedes.
The participant entered her response on a keyboard. She was instructed to sit comfortably and pay as much as possible attention to the stimuli. While perceiving stimuli a white star appeared on the screen, but participants were not obligated to stare at it without moving.
At the beginning of the experimental session participants were instructed on their task. The task was the same in all conditions: they had to state (by pressing two different keys on the keyboard) after each stimulus whether they perceived the stimulus as having an intonation of an affirmative utterance or of an interrogative utterance. 10 trials of practice were randomly played at the very beginning of the experimental session, with one sound for each level of the first condition they were going to be exposed to. The experimental session was divided in two parts given the overall length of the experiment. Participants perceived at the beginning the first and second batteries (conditions) of stimuli, according to the sequence, then they took a break where they could walk and chat a bit with the experimenter. During the pause 2 questionnaires were administered. They pertained to the participant's style to communicate and relate with other individuals. The 2 questionnaires were the Autism Spectrum Quotient (AQ) (Baron-Cohen et al, 2001)

and the Interpersonal Reactivity Index (IRI) (Davis, 1983). Then, the participants participated at the second session where the third, the fourth and again the first battery of stimuli were presented. Each experimental session, pauses between blocks and between conditions included, lasted overall about 2 hours and 30 minutes. Participants were instructed to take breaks between blocks and conditions as long as they needed, therefore in some case the experiment could last about 15 minutes less or longer.

2.6. *Results*

Of the 37 collected participants, 3 were excluded from the analysis because they did not meet the screening criteria. The total number of considered participants in the analysis is 34.

The first step of the analysis consisted of a visual inspection of the data. Curves not showing the expected shape of the logistic function, displaying either a flat contour or a contour shifting in the opposite direction, with respect to the experimental task could indicate that the participant did not correctly perform the task. The Log-Likelihood values (LL) of the fitted curves have been adopted as an indicator for the presence of single fittings that were unlikely to occur. The more an LL value is lower than 0, the less the fitted curve is likely to be fit during the perception and categorization of the presented stimuli. Fitted curves with LL values below the lower than the Tukey's low inner fence of the LL distribution have been considered outliers (Tukey, 1977).

2.6.1. Excluded Participants

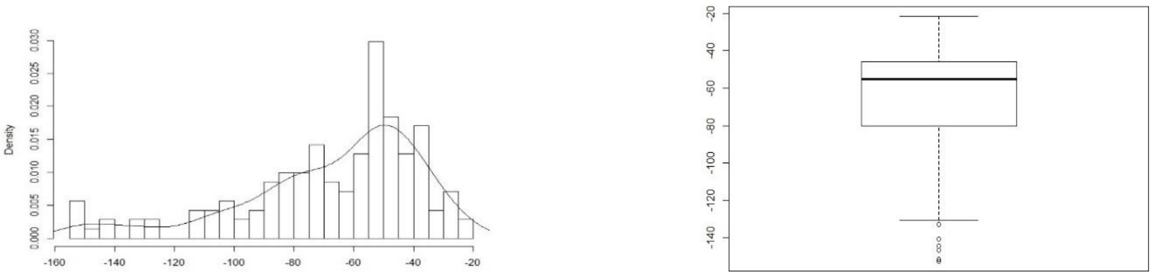


Figure 2.6.1 Distribution and Boxplot of LL values of 34 participants

2.6.1.1. Descriptive statistics and outliers' definition

Descriptive Stats of LL					
Min.value	**1st Quantile**	**Median**	**Mean**	**3rd Quantile**	**Max.value**
-152.49	-82.25	-56.80	-67-60	-46.61	-21.61

Table 2.6.1 Descriptive Stats of LL

Criterion for participants' exclusion – Outliers below the lower Tukey limit		
Higher Limit	Q3 + 1.5* IQR	6.85
Lower Limit	Q1 – 1.5* IQR	-135.71

Table 2.6.2 Calculation of the Tukey Limit

The LL analysis reported 5 participants (nsub: 14, 27, 31, 32, 36) with at least one unlikely psychophysical curve of the 5 generated, one for each experimental condition in each sequence: 1 participant out of 16 did not adequately categorize the humming condition in sequence 1 (occurring in 4th position along the sequence), and 4 participants out of 18 did not adequately categorize the humming condition in sequence 2 (occurring in 1st position in the sequence).

Below, figures from 2.6.2. to 2.6.6. report the graphic representation of responses for each excluded participant.

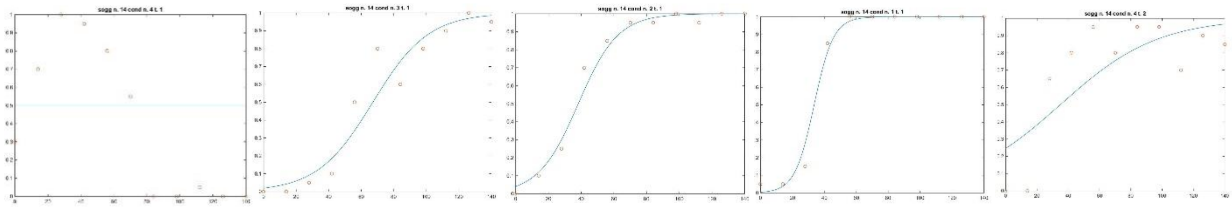


Figure 2.6.2 Excluded nsub14

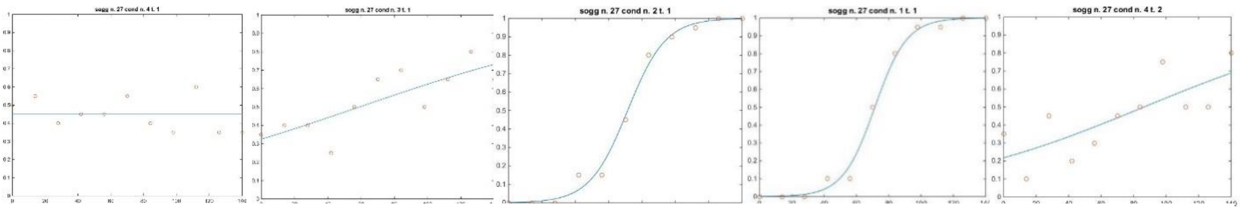


Figure 2.6.3 Excluded nsub27

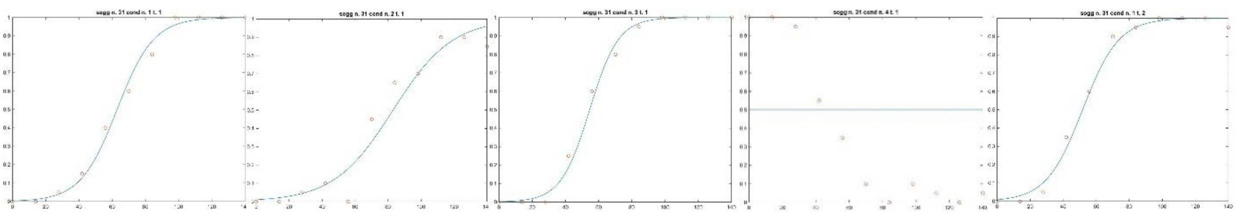


Figure 2.6.4 Excluded nsub31

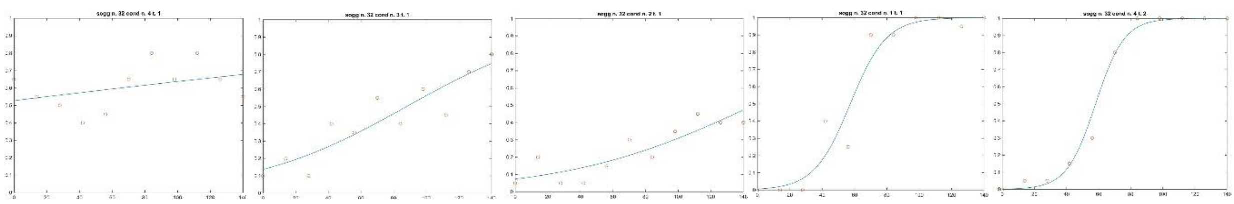


Figure 2.6.5 Excluded nsub32

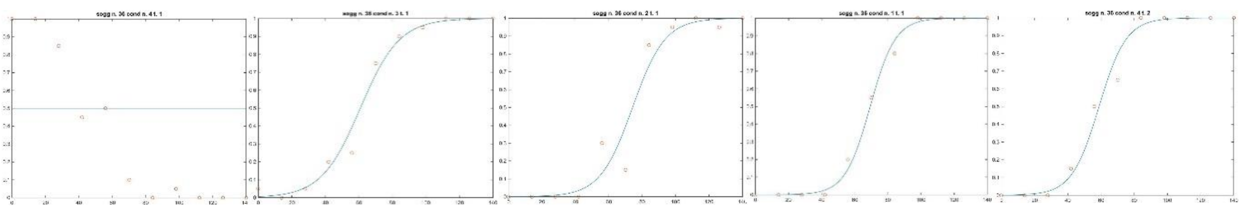


Figure 2.6.6 Excluded nsub36

By a first visual inspection, it is possible to observe that the only condition in which participants displayed to have more difficulty in the categorization process is the humming condition in both sequences. Despite the proportions of unfitted curves in sequence 1 (1/16) and in sequence 2 (4/18) is not significantly different

(binomial test, p = 0.2166), probably due to the limited number of observations, the humming condition in 1st position in sequence 2 has a rate of categorization's failure of the 22%, meaning that one participant out 5 displays difficulties in performing the task.
Below, tables from 2.6.3. to 2.6.6. report descriptive statistics of slope and midpoint values in each condition of in sequence 1 and sequence 2 of the remaining 15 participants in sequence 1 and 14 participants in sequence 2.

2.6.2. Descriptive Statistics

Slope values in Sequence 1 (w_to_hum)							
Order	Conditions	Mean	S.D.	Var.	Min	Max	Obs.
1	word-test	0.0819	0.0328	0.0011	0.0220	0.1330	15
2	pseudoword	0.0845	0.0409	0.0017	0.0320	0.1700	15
3	foreign	0.0741	0.0369	0.0014	0.0190	0.1500	15
4	humming	0.0898	0.0307	0.0009	0.0510	0.1590	15
5	word-retest	0.0778	0.0255	0.0007	0.0340	0.1220	15

Table 2.6.3 Slope values in Sequence 1

Slope values in Sequence 2 (hum_to_w)							
Order	Conditions	Mean	S.D.	Var.	Min	Max	Obs.
1	humming-test	0.0888	0.0321	0.0010	0.0290	0.1600	14
2	foreign	0.0629	0.0262	0.0007	0.0240	0.1320	14
3	pseudoword	0.0746	0.0312	0.0010	0.0300	0.1360	14
4	word	0.1042	0.0542	0.0030	0.0540	0.2200	14
5	humming-retest	0.1046	0.0358	0.0013	0.0630	0.2040	14

Table 2.6.4 Slope values in Sequence 2

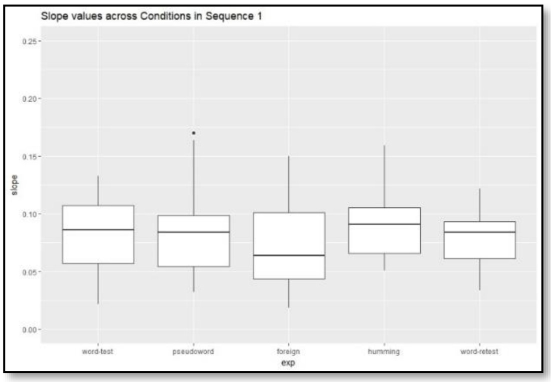


Figure 2.6.8 Slope values in Sequence 1 and Sequence 2

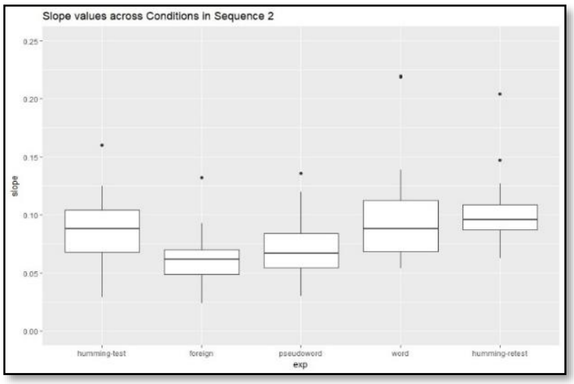


Figure 2.6.7 Slope values in Sequence 1 and 2

Midpoint values in Sequence 1 (w_to_hum)							
Order	Conditions	Mean	S.D.	Var.	Min	Max	Obs.
1	word-test	61.0238	11.4527	131.1664	41.8380	85.3380	15
2	pseudoword	63.1977	14.1128	199.1711	34.9950	96.6620	15
3	foreign	54.5360	16.8241	283.0521	32.9980	102.9020	15
4	humming	68.7028	7.7887	60.6646	60.1970	88.2460	15
5	word-retest	61.6720	11.9721	143.3315	34.6940	81.9360	15

Table 2.6.5 Midpoint values in Sequence 1

Midpoint values in Sequence 2 (hum_to_w)							
Order	Conditions	Mean	S.D.	Var.	Min	Max	Obs.
1	humming-test	73.3810	14.3960	207.2453	51.0350	111.6460	14
2	foreign	65.5032	15.1414	229.2616	43.3900	93.7210	14
3	pseudoword	67.0004	12.9975	168.9352	50.3990	95.3370	14
4	word	54.9185	12.6761	160.6830	30.2820	78.4130	14
5	humming-retest	65.6728	12.2729	150.6239	45.4390	87.5450	14

Table 2.6.6 Midpoint values in Sequence 2

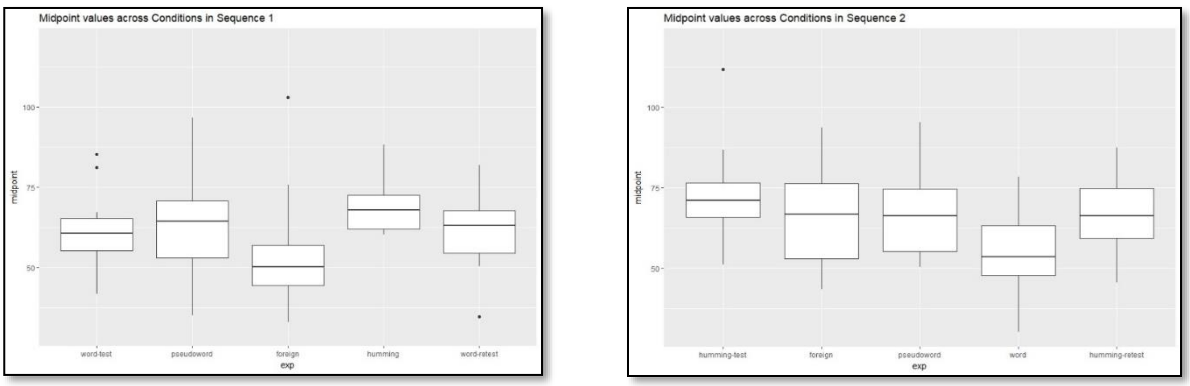


Figure 2.6.9 Midpoint values in Sequence 1 and 2

The inspection of the descriptive statistics highlights that normality, homogeneity of variance (Levene's tests for both slope and midpoint in each sequence are not significant) and sphericity are all respected (Mauchly's tests for both slope and midpoint in each sequence are not significant).

Since the humming condition in the first position in sequence 2 displays the 22% rate of failure in the categorization process, the first analysis addresses the comparison of slope and midpoint values of the word condition in the first position in sequence 1 and of the humming condition in the 1st position of sequence 2.

2.6.3. Word vs. Humming in sequences' first position comparison

Slope values comparison in first condition								
Order	Sequence	Conditions	Mean	S.D.	Var.	Min	Max	Obs.
1	1	word-test	0.0819	0.0328	0.0011	0.0220	0.1330	15
1	2	humming-test	0.0888	0.0321	0.0010	0.0290	0.1600	14

Table 2.6.7 Slope values comparison in first condition

Midpoint values comparison in first condition								
Order	Sequence	Conditions	Mean	S.D.	Var.	Min	Max	Obs.
1	1	word-test	61.0238	11.4527	131.1664	41.8380	85.3380	15
1	2	humming-test	73.3810	14.3960	207.2453	51.0350	111.6460	14

Table 2.6.8 Midpoint values comparison in first condition

Results of the between subjects t-test of slope values in word condition, compared to slope values in humming condition display that the slope is not significantly different in the two conditions (t = -0.5734, df = 26.93, p = 0.5711, C.I = (-0.0317, 0.0178)). Results of the between subjects t-test of midpoint values in word and humming condition highlight a significant difference (t = -2.5465, df = 24.85, p = 0.0175, C.I. = -22.3545, -2.3599)), showing that the humming condition in 1st position has a mean midpoint value 12 Hz higher than the word condition.

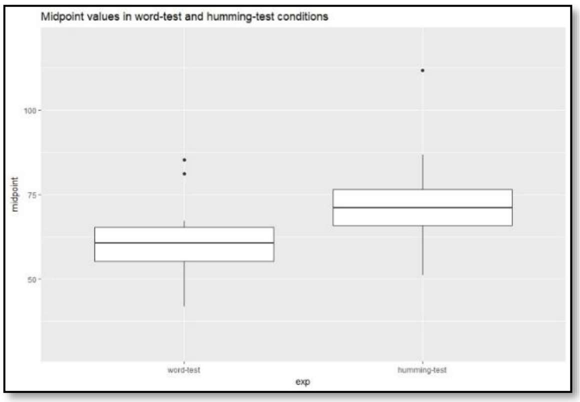


Figure 2.6.10 Midpoint values in word and humming test conditions

Since the interest of this study is to observe the processing differences generated by the two reversed orders of conditions in two sequences, and since the humming condition in 1st position showed to be more difficult to be categorized, and to have a significantly different category boundary between the statement and question categories, the following step of the analysis investigates whether along the two sequences, considered as two factors with 4 levels, values of slope and midpoint in each condition are respectively correlated. A further interest is to observe whether the correlation along conditions is similar across sequences or different, because of the reversed order.

The retest condition is temporarily excluded from the analysis and will be separately investigated afterward. In the following paragraphs, Pearson's Product-Moment correlation tests of the 6 possible conditions' comparisons in each sequence are presented, p. values are FDR adjusted.

2.6.4. Correlations across conditions in the two sequences

Correlations of slope values across conditions Sequence 1

Pearson's Product-Moment Correlation Test of Slope values in Sequence 1							
Contrast		t	df	Corr.value	p.value	FDR adjusted p.value	Confidence Interval
1	word-test pseudoword	4.0354	13	0.7457	0.0014	0.0043	0.3778 0.9103
2	word-test foreign	4.2986	13	0.7662	0.0009	0.0043	0.4180 0.9181
3	word-test humming	2.4337	13	0.5595	0.0301	0.0386	0.0662 0.8330
4	Pseudoword foreign	2.3123	13	0.5399	0.0378	0.0386	0.0381 0.8242
5	Pseudoword humming	2.3005	13	0.5379	0.0386	0.0386	0.0354 0.8233
6	Foreign humming	2.5458	13	0.5768	0.0244	0.0386	0.0916 0.8407

Table 2.6.9 Correlation of slope values in Sequence 1

Correlations of slope values across conditions Sequence 2

Pearson's Product-Moment Correlation Test of Slope values in Sequence 2							
Contrast		t	df	Corr.value	p.value	FDR adjusted p.value	Confidence Interval
1	humming-test foreign	0.0131	12	0.0038	0.9898	0.9898	-0.5279 0.5333
2	humming-test pseudoword	1.3960	12	0.3738	0.1880	0.2821	-0.1957 0.7547
3	humming-test word	0.7715	12	0.2174	0.4554	0.5464	-0.3540 0.6706
4	foreign pseudoword	1.8539	12	0.4719	0.0885	0.1770	-0.0783 0.8017
5	foreign word	5.1291	12	0.8288	0.0003	0.0015	0.5321 0.9442
6	pseudoword word	3.1380	12	0.6714	0.0086	0.0257	0.2187 0.8863

Table 2.6.10 Correlation of slope values in Sequence 2

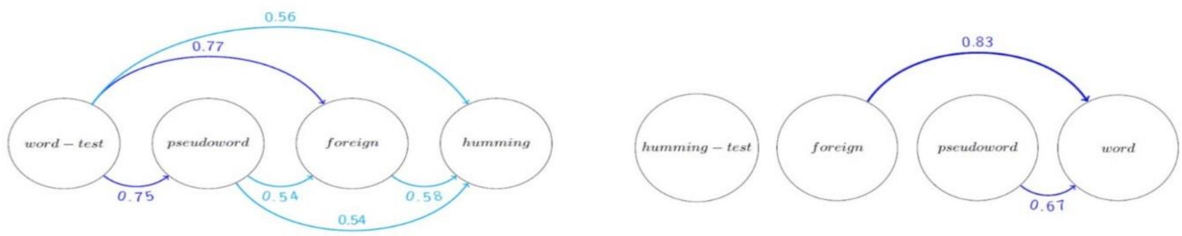


Figure 2.6.11 Correlation of slope values in Sequence 1 and 2

Correlations of midpoint values across conditions Sequence 1

Pearson's Product-Moment Correlation Test of Midpoint values in Sequence 1							
Contrast		t	df	Corr.value	p.value	FDR adjusted p.value	Confidence Interval
1	word-test pseudoword	4.2948	13	0.7660	0.0009	0.0052	0.4174 0.9180
2	word-test foreign	2.2924	13	0.5365	0.0392	0.0471	0.0335 0.8227
3	word-test humming	1.7925	13	0.4452	0.0964	0.0964	-0.0869 0.7796
4	Pseudoword foreign	3.4536	13	0.6917	0.0043	0.0120	0.2780 0.8890
5	Pseudoword humming	3.2769	13	0.6726	0.0060	0.0120	0.2446 0.8812
6	Foreign humming	2.5233	13	0.5734	0.0255	0.0382	0.0865 0.8392

Table 2.6.11Correlation of midpoint values in Sequence 1

Correlations of midpoint values across conditions Sequence 2

Pearson's Product-Moment Correlation Test of Midpoint values in Sequence 2							
Contrast		t	df	Corr.value	p.value	FDR adjusted p.value	Confidence Interval
1	humming-test foreign	3.0875	12	0.6654	0.0094	0.0319	2.083e-01 0.8839
2	humming-test pseudoword	2.1683	12	0.5306	0.0510	0.0764	-2.241e-05 0.8280
3	humming-test word	1.0163	12	0.2815	0.3296	0.3296	-2.928e-01 0.7066
4	foreign pseudoword	3.0216	12	0.6573	0.0106	0.0319	1.946e-01 0.8807
5	foreign word	1.2907	12	0.3492	0.2211	0.2653	-2.227e-01 0.7422
6	pseudoword word	2.6613	12	0.6092	0.0208	0.0415	1.162e-01 0.8614

Table 2.6.12 Correlation of midpoint values in Sequence 2

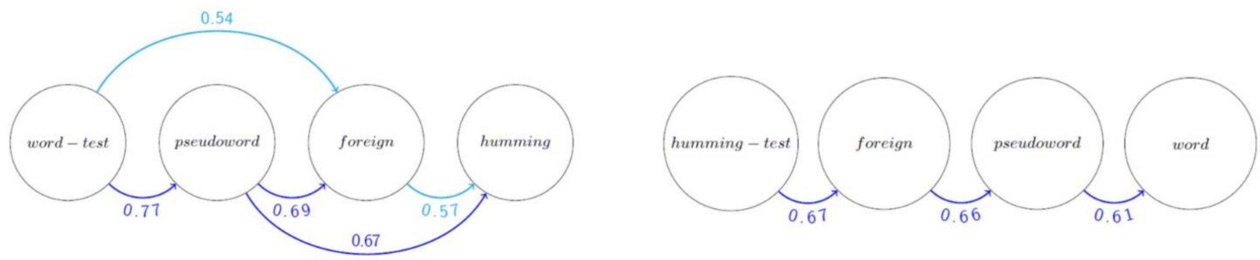


Figure 2.6.12 Correlation of midpoint values in Sequence 1 and 2

In sequence 1 (word, pseudoword, foreign, humming) slope values display to be highly intercorrelated among conditions, in particular between the initial word condition and the other 2 carrying linguistic segmental information (pseudoword, r = 0.75; foreign, r = 0.77). Slope values in sequence 2, on the other hand, correlate only between conditions with linguistic segmental information and the word condition in 4th position (foreign, r = 0.83; pseudoword, r = 0.67).
Midpoint values in sequence 1 show also to be highly intercorrelated, except for the comparison between the word condition in 1st position and the humming condition. The mean midpoint value in the humming condition in 4th position within sequence 1 is 7 Hz higher than in the word condition in 1st position.
The midpoint values in sequence 2 show to be highly intercorrelated only with the values in the adjacent condition.
The correlation analysis is a first piece of evidence that the two reversed orders in the two conditions generate a different processing when performing a linguistic categorization task of pitch with increasing (sequence 2) or degrading (sequence 1) segmental information in each condition.

The descriptive analysis and correlational analysis show that when the categorization task is performed firstly with the full linguistic information available in the stimuli, as in the word condition in 1st position in sequence 1, the properties of the categorization process appear to be transferred along the sequence and used in the categorization task of the following conditions. Except for the position of the category boundary in the humming condition, both the slope and midpoint highly correlate across conditions. In sequence 2, slope values correlate only between the word in the condition in 4th position and the antecedent linguistic conditions, and the midpoint is influenced only by the antecedent condition. This difference between the sequences is also observable looking at the boxplots. In sequence 1, conditions' mean values of both slope and midpoint appear to be within a similar range across conditions; on the contrary in sequence 2 it is possible to observe looking at mean values that the slope parameter appears to increase, and the midpoint appears to decrease. I investigate therefore with Linear Mixed Models (LMEMs) the presence of a group and/or individual trend in both sequences. LMEMs are constructed independently in each sequence considering slope and midpoint independently as single dependent variables. The 4 conditions in each sequence are transformed in a single predictor called expseq in sequence 1, and expseq2 in sequence 2 which is composed of 4 ordered numeric levels going from 0 (condition 1) to 3 (condition 4). Intercepts are centered on level 0.
In expseq the level 0 corresponds to the word-test condition, in expseq2 the level 0 corresponds to the humming-test condition.

2.6.5. Linear Mixed-effects Models (LMEMs) with 4 conditions

For each sequence (1, 2) and each dependent variable (slope, midpoint) the two maximal models are compared. They comprehend the fixed effect of the predictor expseq/expseq2, the random effect of individual variability of intercept-values (1|nsub) and the random effect of individual variability of the slope of the linear regression model across conditions or (expseq2|nsub).
The following models are considered:
- fm3, the model comprehending fixed and correlated random effects (~ expseq + (expseq|nsub));
- fm2, the model comprehending fixed and where random effects are left free not to correlate (~ expseq + (1|nsub) + (0 + expseq|nsub)):
- fm1, the model with only the fixed effects and random intercept (~ expseq + (1|nsub)).

Then, the cases in which the fixed effect of the sequence's predictor does not significantly explain any variance are considered, as well as the models that consider the random effects of the participants.
The following models are thus added to the analysis:

- m3, the model considering only correlated random intercept and slope of the linear model (~ (expseq|nsub));
- m2, the model that considers the case in which random effects are not correlated (~ (1|nsub) + (0 + expseq|nsub));
- m0, the basic model comprehending only the random intercepts (~ (1|nsub)).

The resulting Mixed Models are compared starting from the more complex ones comprehending both fixed and random effects and then moving to the simplest ones. In case two models significantly fit the data, the simplest model is preferred. Figure 2.6.13 below represents the order of comparison.

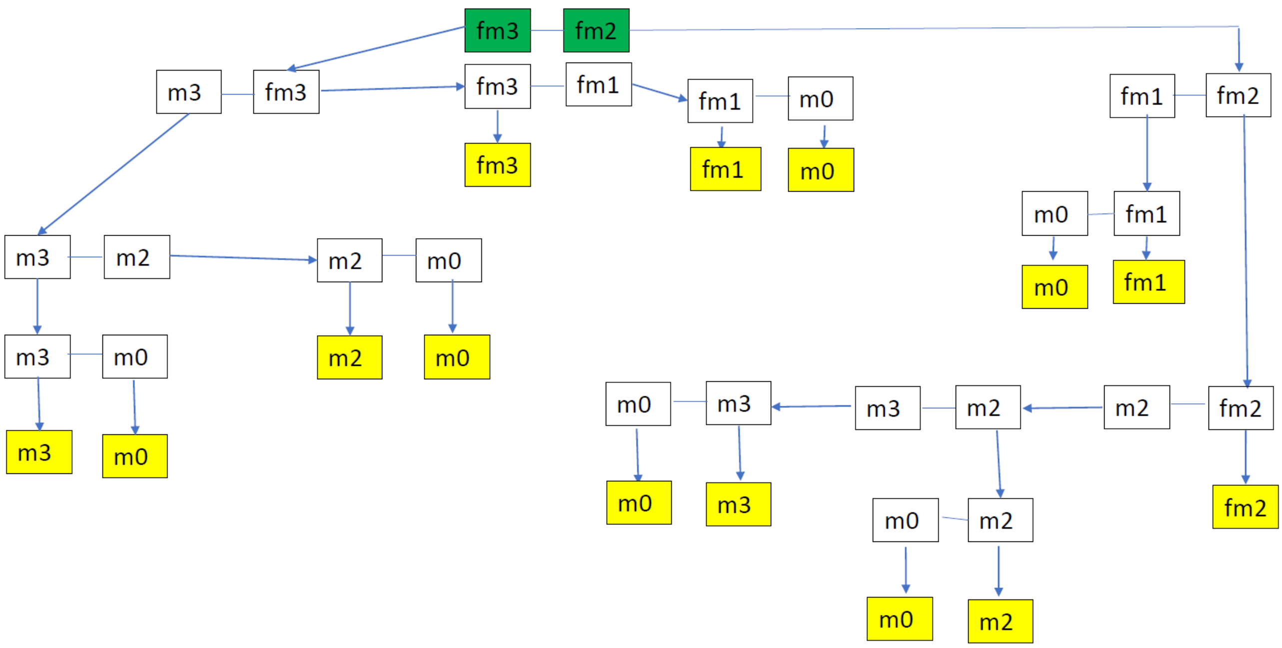


Figure 2.6.13 Hierarchical analysis of AoV LMEMs

2.6.4.1. <u>LMEMs of slope and midpoint values in sequence 1 – 4 conditions</u>

The ANOVA model comparison of slope values in sequence 1 does not display any significant effect, in any model comparison, and m0 is therefore the best model describing the data (beta0 = 0.0826, SE = 0.0077, t = 10.68, p = 4.1e-08). Slope values are not significantly different across conditions, the levels of the sequences do not systematically vary slope values, and no variation of the intercept is related to beta values along the conditions. The midpoint dependent variable displays the same behavior as the slope dependent variable. ANOVA model comparison of midpoint values indicates m0 as best model (beta0 = 61.865, SE = 2.765, t = 22.38, p = 2.33e-12)

2.6.4.2. LMEMs of slope values in sequence 2 – 4 conditions

The ANOVA model comparison of slope values in sequence 2 does not display any significant difference between fm2 (uncorrelated random effect) and fm3 (correlated random effects).
Therefore, fm2 is compared with fm1 resulting the best linear fitting.

Model	Df	AIC	BIC	Chisq	p.value
Fixed Effect and random intercept (fm1)	4	-203.38	-195.28		
Fixed Effect and uncorrelated random intercept and slope effect (fm2)	5	-208.08	-197.96	6.6975	0.0097

Table 2.6.13 LMEMs of slope values in Sequence 2 - selecting fm2

Results highlight a non-significant fixed effect of expseq2 on slope values (beta = 0.0058, SE = 0.0051, t = 1.144, p = 0.268) and significant effect of the intercept (beta0 = 0.0739, SE = 0.0068, t = 10.894, p = 3.91e-09).
Since the sequence predictor is not significant, fm2 is further compared with the equivalent model that does not consider the fixed effect predictor.

Model	Df	AIC	BIC	Chisq	p.value
uncorrelated random intercept and random slope effect (m2)	4	-208.76	-200.66		
Fixed Effect and uncorrelated random intercept and slope effect (fm2)	5	-208.08	-197.96	1.3202	0.2506

Table 2.6.14 LMEMs of slope values in Sequence 2 - selecting m2 over fm2

m2 is chosen and further compared with m3 (uncorrelated random effects without the sequence predictor) and with the simplest model m0.

Model	Df	AIC	BIC	Chisq	p.value
uncorrelated random intercept and random slope effect (m2)	4	-208.76	-200.66		
correlated random intercept and slope effect (m3)	5	-207.29	-197.16	0.528	0.4674

Table 2.6.15 LMEMs of slope values in Sequence 2 - selecting m2 over m3

Model	Df	AIC	BIC	Chisq	p.value
Random intercept (m0)	3	-203.17	-197.10		
uncorrelated random intercept and random slope effect (m2)	4	-208.76	-200.66	7.5883	0.0059

Table 2.6.16 LMEMs of slope values in Sequence 2 - selecting m2 over m0

m2 is chosen as best fit, since it is preferable over m3 and m0 (beta0 = 0.0779, SE = 0.0058, t = 12.42, p = 8 .6e-09).

2.6.4.3. LMEMs of midpoint values in sequence 2 – 4 conditions

The ANOVA model comparison of midpoint values in sequence 2 does not display any significant difference between fm2 (uncorrelated random effect) and fm3 (correlated random effects), and in second instance does not display a significant effect when fm2(uncorrelated random effect) and fm1 (fixed effect with random intercept) are compared. There is a significant main effect of expseq2 when fm1 is compared to m0 and therefore the choice is fm1 (beta = -5.389, SE = 1.188, t = -4.537, p = 4.91e-05; beta0 = 73.284, SE = 3.447, t = 21.258, p < 2e-16).

.

Model	Df	AIC	BIC	Chisq	p.value
Random intercept (m0)	3	459.51	465.59		
Fixed Effect and random intercept (fm1)	4	444.43	452.53	17.086	3.57e-05

Table 2.6.17 LMEMs of midpoint values in Sequence 2 - selecting fm1

Fixed Effect and random intercept Model (fm1)				
Fixed Effects:	**Estimate**	**SE**	**t**	**p.value**
Intercept(beta0) Individual variability	73.284	3.447	21.258	<2e-16
expseq2(beta)	-5.389	1.188	-4.537	4.91e-05

Table 2.6.18 LMEMs of midpoint values in Sequence 2 - fm1 model

The analyses with 4 conditions display that in sequence 1 the order of conditions plays no effect in slope and midpoint variation. High variability is present also when categorizing a fully linguistic stimulus (word) in the degree of categorical perception displayed by participants (slope values) and in the position of the categories boundaries (midpoint). Nonetheless, the correlational analysis of mean values across conditions highlights that in sequence 1 both slope and midpoint (except for the humming condition) are highly correlated. Together these data suggest that in sequence 1 all participants adopt a similar strategy along the conditions, with mean values non-significantly different in each condition, but that that strategy can be significantly different across subjects. This information, to start, is interesting for the debate on the nature of pitch perception and the scientific question of whether categorical perception is present or not. “Pure” CP might be hard to find if the criteria with which the stimuli are categorized in fully linguistic stimuli significantly vary across subjects, and if also the category boundaries vary to a great extent.

In sequence 2, the reversed order of the sequence (from lowly linguistic to highly linguistic segmental information) plays a role, but differently when comparing slope and midpoint. Slope values show to be best explained by a model (m2) that considers a random intercept (slope in word condition) and an uncorrelated slope of the model (the increase of slope along conditions of each participant). This means that the sequence creates an effect of perceptual training systematically increasing or decreasing the slope value (an index of certainty of the participant whether the stimulus belong to one of the two categories) but that the direction and the degree of variation varies across subjects, also considering that slope values in the first presented condition (humming-test) vary across subjects, too.

The midpoint trend in sequence 2 (fm1) is the one showing the tightest connection with the sequence, since a main fixed effect of the sequence emerges, indicating that the midpoint systematically decreases along conditions.

This is also supported by the correlational analysis which highlights how the midpoint in each condition, at the group level, correlates with the adjacent condition only.
The 2 sequences have been designed with 4 conditions with the aim to find a continuous linear trend in the way pitch is perceived when hardly recognizable segmental content is aligned to pitch and with the aim to observe the effect in two opposite directions when considering the two experimental orders.
Nonetheless, this analysis shows on one side that when the sequence begins with recognizable segmental information aligned to pitch the criterion with which is categorized is kept constant along the sequence, but it depends from individual variability more than from the order of the stimuli. On the other one, the analysis shows that when the sequence begins with stimuli with hardly recognizable segmental information, the starting variability significantly increases since one participant out 5 is not successful in computing the task, the strategy with which the task is performed is shared only by linguistic conditions (correlational analysis of slope in sequence 2), the humming stimuli appear to be processed in a separated way, and the way the sequence impacts on categorization's performances is different among participants. The position of category boundaries, nonetheless, shows to behave as expected and to averagely decrease, in function of task's complexity, from humming to word.
These results together with the visual inspection of the data suggest that the processing of pitch aligned to humming segmental material might not stand in a linear relationship with the other conditions. In the next analysis, the same sequences and possible models deprived of the humming condition, are considered.
The numerical factor expseq shifts in the following analysis from 0 to 2, in sequence 1, and expseq2 is recoded with the level 0 corresponding to the foreign condition in sequence 2, shifting from 0 to 2 from the foreign to word condition. The main interest is to observe whether slope and midpoint trends are still independent from the sequence predictor in sequence 1, or on the contrary a stable relationship emerges. Secondly, the aim is to observe the presence of a perceptual learning process linearly increasing slope values, and linearly decreasing midpoint values in sequence 2, without the variance that the humming condition might have introduced in the previous analysis.

2.6.5. LMEMs of only linguistic segmental information (3 conditions)

2.6.5.1. LMEMs of slope values in sequence 1 – word, pseudoword, foreign

The ANOVA model comparison of slope values in sequence 1 does not display any significant effect, in any model comparison, and m0 is therefore the best model describing the data (beta0 = 0.0802, SE = 0.0084, t = 9.502, p = 1.75e-07).

2.6.5.2. LMEMs of midpoint values in sequence 1 – word, pseudoword, foreign

The best model describing midpoint values in sequence 1 when only linguistic conditions (3) are considered is fm1 which shows the tendency to be better than m0 (beta= -3.244, SE = 1.649, t = -1.967, p = 0.0588; beta0 = 62.830, SE = 3.614, t = 17.385, p = 3.48e-14). The absence of the humming condition lets the fixed effect of the sequence emerge.

Model	Df	AIC	BIC	Chisq	p.value
Random intercept (m0)	3	359.60	365.02		
Fixed Effect and random intercept (fm1)	4	357.84	365.06	3.7584	0.0525

Table 2.6.19 LMEMs of midpoint values in Sequence 1 - selecting fm1

2.6.5.3. LMEMs of slope values in sequence 2 – foreign, pseudoword, word

Without the humming condition, slope values in sequence 2 are best described by the most comprehensive model fm3, predicting the fixed effect of the sequence, and correlated random intercept and slope (beta= 0.0207 SE = 0.0055, t = 3.752, p = 0.0020; beta0= 0.0599 SE = 0.0066, t = 9.014, p = 1.44e-07).

Model	Df	AIC	BIC	Chisq	p.value
Fixed Effect and uncorrelated random intercept and slope effect (fm2)	5	-163.40	-154.72		
Fixed Effect and correlated random intercept and slope effect (fm3)	6	-167.16	-156.74	5.759	0.0164

Table 2.6.20 LMEMs of slope values in Sequence 2 - selecting fm1 over fm2

Model	Df	AIC	BIC	Chisq	p.value
Fixed Effect and random intercept (fm1)	4	-160.69	-153.74		
Fixed Effect and correlated random intercept and slope effect (fm3)	6	-167.16	-156.74	10.475	0.0053

Table 2.6.21 Table 2.6.20 LMEMs of slope values in Sequence 2 - selecting fm3 over fm1

2.6.5.4. LMEMs of midpoint values in sequence 2 – foreign, pseudoword, word

Analysis of midpoint trend in sequence 2, without the presence of the humming condition in the model, highlights that fm1, comprehending the fixed effect of the sequence and the random effect of the intercept of each participant, is still the best model describing the data (beta= -5.292 SE = 1.891, t = -2.799, p = 0.0093; beta0= 67.766 SE = 3.567, t = 19.001, p = 9.84e-16).

Model	Df	AIC	BIC	Chisq	p.value
Random intercept (m0)	3	342.55	347.76		
Fixed Effect and random intercept (fm1)	4	337.41	344.37	7.1353	0.0076

Table 2.6.22 Table 2.6.20 LMEMs of midpoint values in Sequence 2 - selecting fm1

2.6.5.5. Summary of LMEMs' analysis

Overall the LMEMs' analysis shows that the presence or absence of the humming condition plays a role in the models' selection. When the categorization process is based on a pragmatic and syntactical function, such as sentence modality, it is possible that pitch aligned to segmental information that carries poorly recognizable linguistic information, such as humming, is differently processed compared to cases in which it is aligned to fully linguistic stimuli. The degree of recognizability of segmental information becomes a determinant predictor of the slope parameter in sequence 2 when the humming condition is removed. Slope trends along conditions in sequence 2, when only the linguistic conditions are considered, show to be highly dependent from the recognizability of segments. Precisely, perception becomes increasingly categorical when shifting from the perception of segments of an unrecognizable language, to the perception of stimuli carrying meaning in native language. In second instance, without the humming condition, the category boundaries are set reflecting a process of perceptual training based on extrinsic references across trials. This emerges more consistently in sequence 2 where the effect of the sequence is stable also in the presence of the humming condition, but also in sequence 1, in the opposite direction from the one expected (it slightly diminishes, 3 Hz, instead of increasing in function of difficulty in recognizability of segments) where without the humming condition shows to be related to the sequence.

Interestingly, the humming condition does not affect the categorization process in sequence 1 where the determinant fact is the criterion adopted by participants when perceiving recognizable speech in their native language. Each participant with an individual strategy can transfer that criterion over the other conditions independently from the recognizability of segments. Individual variability is highly emerging in the categorization of the foreign condition's stimuli. Contrary to expectations, the slope values do not decrease along sequence 1, but they are averagely kept constant, with and without humming condition. This, together with unexpected midpoint behavior in sequence 1, highlights that also the sequence 1 instantiates a perceptual learning process, and participants take advantage of it along the conditions, improving their performances. Despite the fact, that both sequences generate perceptual training, the role of recognizability of segments is still determinant since the two sequences generate significantly different models.

Moreover, results highlight that slope and midpoint are sensitive to different properties of speech and their values are differently modeled in the two sequences. Results suggest that, when the first condition is sufficiently rich of information to match top-down activation's expectations (word), both slope and midpoint reflect a linguistic categorization, and that same strategy remains top-down activated also in other conditions until the humming condition is presented (sequence 1). In that case, the strategy of categorizing the humming condition in position 4 (slope) correlates with the linguistic one, but the midpoint position shifts to the middle value of the available range of frequencies (in this case, at about 70 Hz). This is observable in LMEMs' of sequence 1 with and without humming where the trend inverts direction (increase vs. decrease) when the humming condition is removed.

In the reversed order (sequence 2), the categorization of humming in the first position analogously places the midpoint at about 70 Hz, but the strategy with which the stimuli are categorized (slope) is independent of the other linguistic conditions.

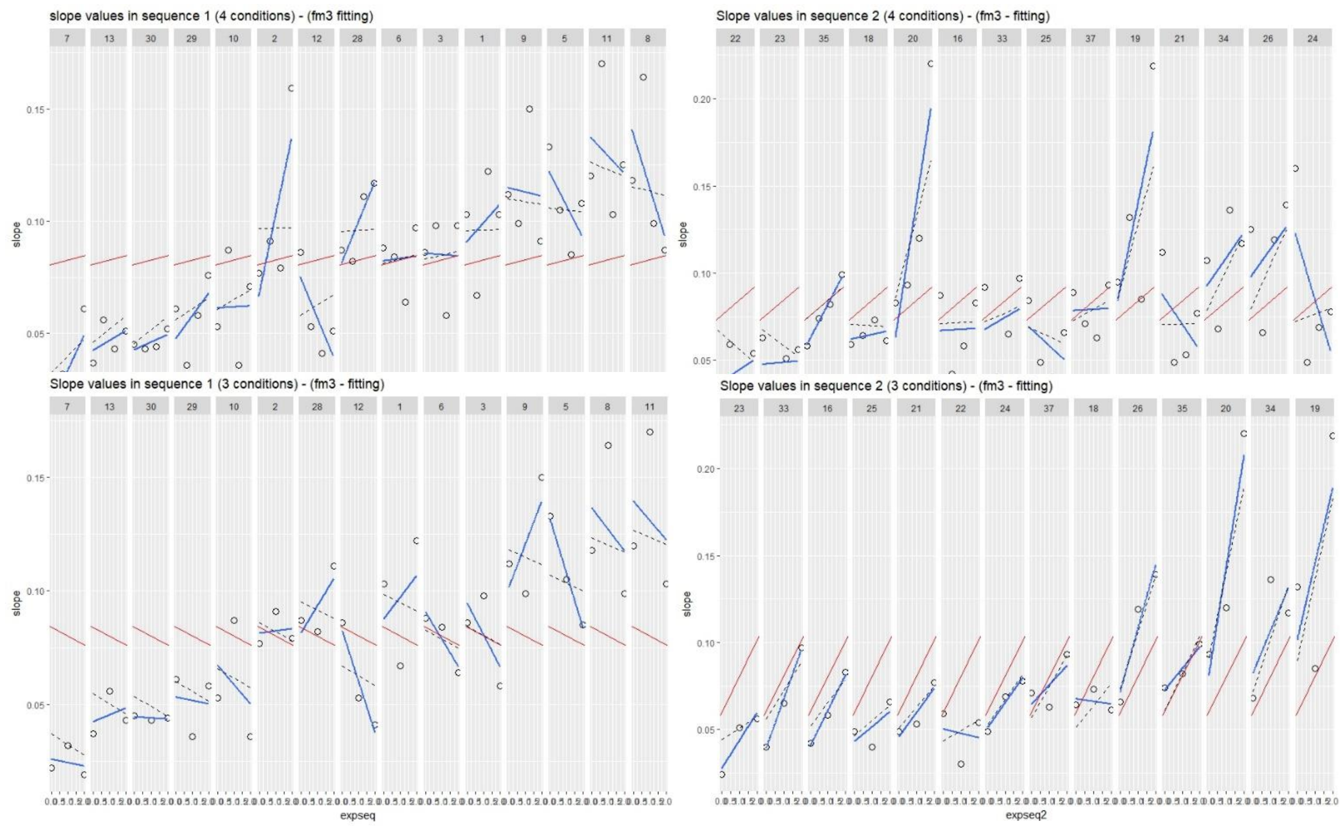
This highlights that the midpoint is a relevant cue of bottom-up extrinsic references processing across trials, therefore of the specific categorization process opposed to identification, that can be used to overcome a strict speech vs. non-speech's signal-dichotomy; on the other hand, these results highlight that the slope is a relevant parameter to separately investigate two different perception processes, matching different top-down expectations, and therefore setting the system to be prepared for decoding speech vs. non-speech material, and impairing the system when stimuli lack the expected properties (recognizability of linguistic segments).

The full analysis is summarized in the following table and plot of longitudinal effects. All plots consider fm3 (the full model) when considering the LMEMs' predictions to allow for an easier comparison between sequences.

Comparison of LMEMs selection including and excluding the humming condition			
With the humming condition			
Sequence 1		Sequence 2	
slope	midpoint	Slope	midpoint
m0	m0	m2	fm1
Without the humming condition			
Sequence 1		Sequence 2	
slope	midpoint	Slope	midpoint
m0	fm1	fm3	fm1

Table 2.6.23 Overall LMEMs selection

Longitudinal trends of slope values in sequence 1 and sequence 2, with and without humming conditions



= value in each condition
= population mean = lm regression within subjects = fitting of fm3 LMEM's predicted values

Figure 2.6.14 Longitudinal trends of slope values in sequence 1 and sequence 2, with and without humming conditions

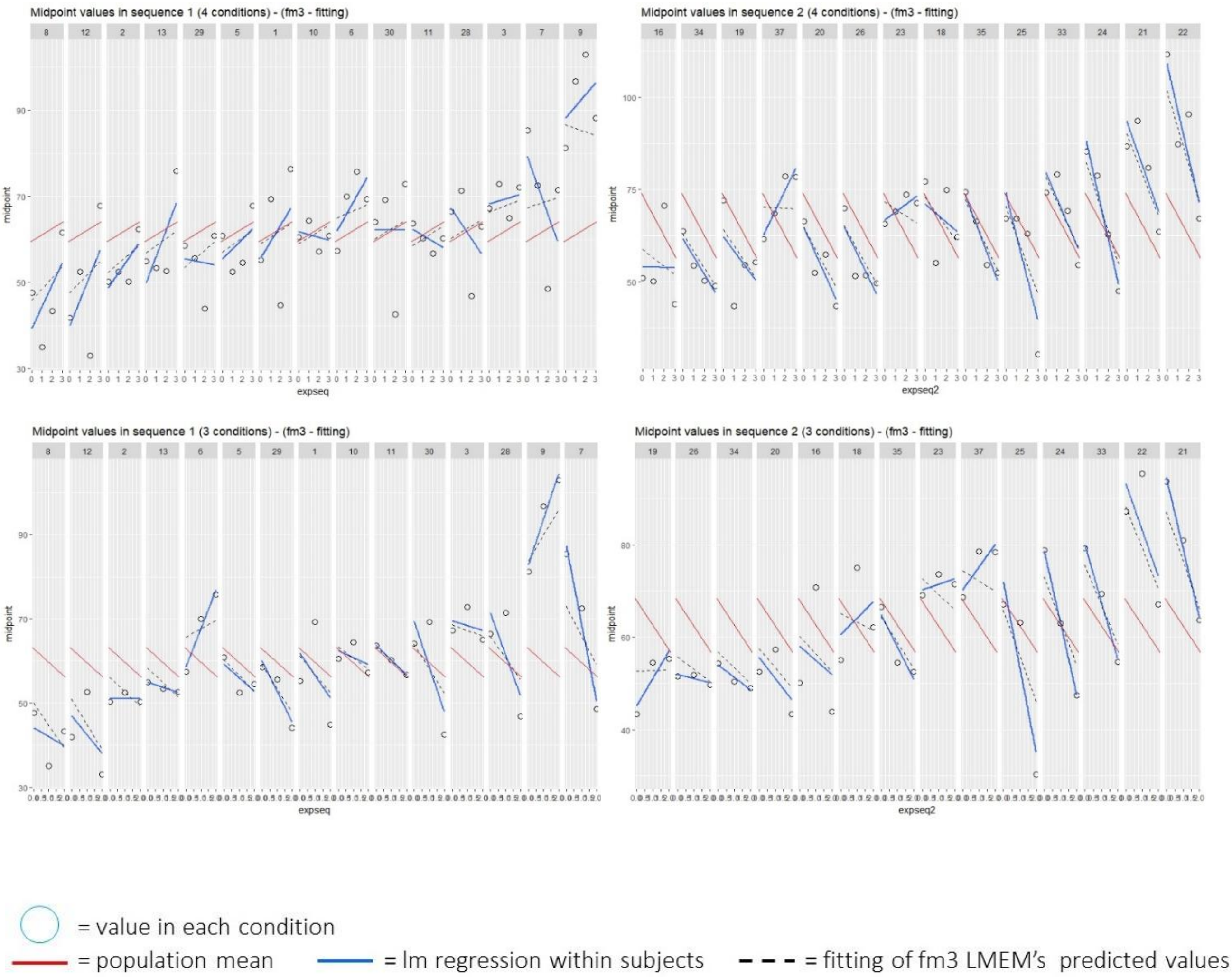


Figure 2.6.15 Longitudinal trends of midpoint values in sequence 1 and sequence 2, with and without humming conditions

The last step of the analysis considers the effect of the sequence on the retest condition. The aim is to observe whether the same exact task in the two sequences is consistently categorized with the same strategy as in the test condition, or significantly different. This is an indication about the attention of the participants along the experiment (which was considerably long) and an indication of whether if the retest phase impacts similarly or differently on the word and humming conditions.

2.6.6. Test – Retest phase's Results

Word-test and word-retest conditions are compared in sequence 1 and the humming-test and humming-retest condition in sequence 2. The Table 2.6.24 reports group-level descriptive stats (mean, S.D, Var and number of observations) for slope and midpoint in the test and retest conditions and group – level differences in the average slope and midpoint values.

Descriptive Statistics

Parameter	word (sequence 1)							
	Test				Retest			
	Mean	S.D.	Var.	Obs.	Mean	S.D.	Var.	Obs.
Slope	0.0819	0.0328	0.0011	15	0.0778	0.0255	0.0007	15

Parameter	humming (sequence 2)							
	Test				Retest			
	Mean	S.D.	Var.	Obs.	Mean	S.D.	Var.	Obs.
Slope	0.0888	0.0321	0.0011	14	0.1046	0.0356	0.0013	14

Parameter	word (sequence 1)							
	Test				Retest			
	Mean	S.D.	Var.	Obs.	Mean	S.D.	Var.	Obs.
Midpoint	61.0238	11.4528	131.1664	15	61.6720	11.972	143.3315	15

Parameter	humming (sequence 2)							
	Test				Retest			
	Mean	S.D.	Var.	Obs.	Mean	S.D.	Var.	Obs.
Midpoint	73.3810	14.3961	207.2453	14	65.6728	12.273	150.6239	14

Table 2.6.24 Descriptive Statistics of Test-Retest comparison

2.6.6.1. Correlations

Table 2.6.25 report the correlation tests for each parameter in both sequences.

Pearson's product-moment correlation	Word Test – Word Retest				Humming Test – Humming Retest			
	Effect	p-value	df	95% C.I.	Effect	p-value	df	95% C.I.
Slope	0.7681	0.0008	13	0.4218 0.9188	0.1828	0.5316	12	-0.3851 0.6503
Midpoint	0.3068	0.2660	13	- 0.2438 0.7078	0.4085	0.1470	12	-0.1558 0.7718

Table 2.6.25 Correlation of slope and midpoint values in Test-Retest comparison

The only significant correlation concerns the slope parameter in sequence 1 and is very strong (r=0.7681, p = 0.0008, and C.I excluding the value of 0)

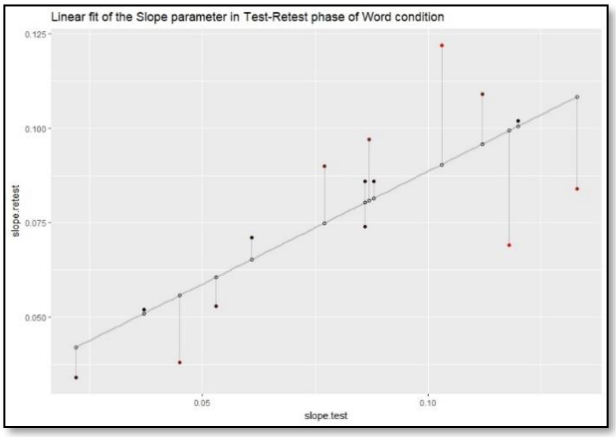
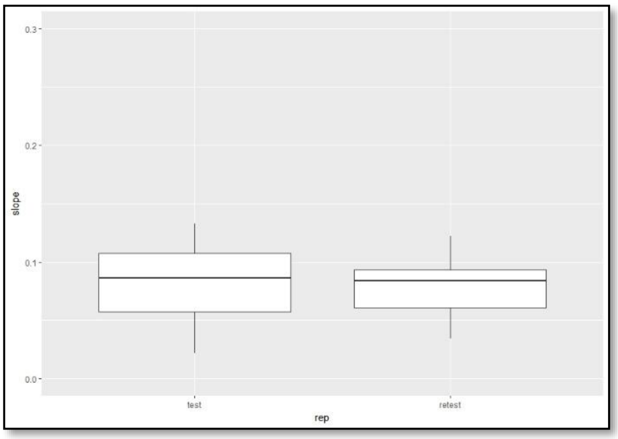


Figure 2.6.16 Correlation of slope values in Test-Retest comparison

2.6.6.2. Mean differences in test-retest phases

In the following sections the possibility of significant differences in mean values for both parameters is explored.

Boxplot of the Slope parameter in Test-Retest phase in Word condition



Boxplot of the Slope parameter in Test-Retest phase in Humming condition

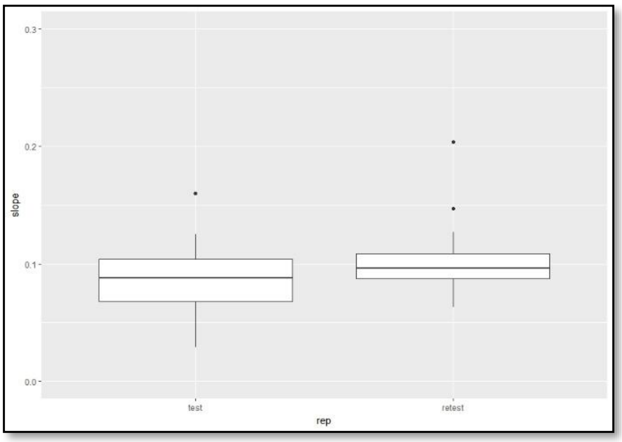
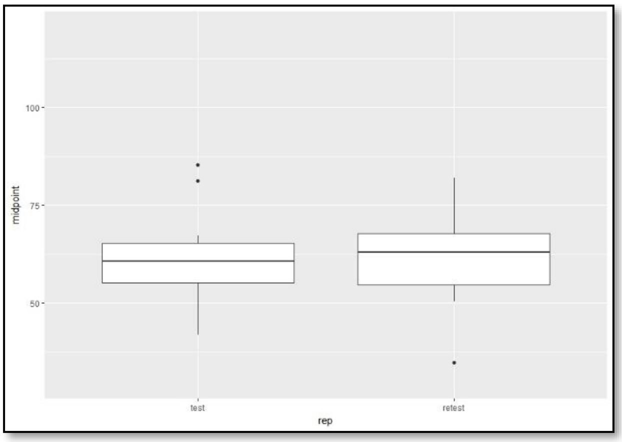


Figure 2.6.17 Slope in word and humming conditions in Test-Retest comparisons

Boxplot of the Midpoint parameter in Test-Retest phase in Word condition



Boxplot of the Midpoint parameter in Test-Retest phase in Humming condition

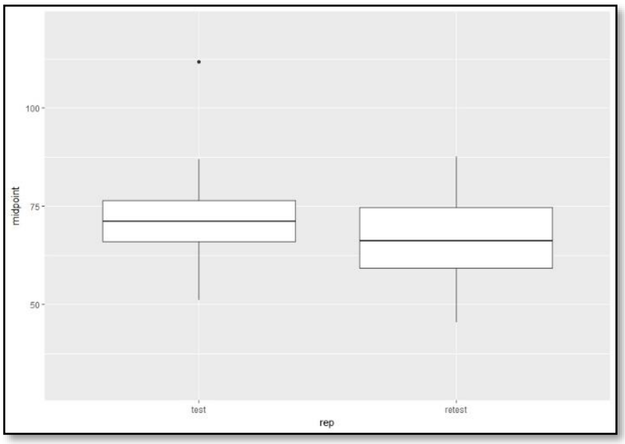


Figure 2.6.18 Midpoint in word and humming conditions in Test-Retest comparisons

2.6.6.3. T-Tests

Pairwise T-Test of Test and Retest Conditions								
Slope	Word Test – Word Retest				Humming Test – Humming Retest			
	t	p.value	df	C.I.	t	p.value	df	C.I.
	0.7495	0.4660	14	-0.0076 0.0157	-1.3562	0.1981	13	-0.0409 0.0094
Midpoint	Word Test – Word Retest				Humming Test – Humming Retest			
	t	p.value	df	C.I.	t	p.value	df	C.I.
	-0.1820	0.8582	14	-8.2888 6.99238	1.9738	0.0700	13	-0.7286 16.1450

Table 2.6.26 Slope and midpoint t-test in Test-Retest comparison

Contrasting the two parameters in test-retest phases in the two sequences, as they were a test-retest comparison of the same condition in a 2 levels repeated-measure design, the parameter midpoint shows a trend to be significantly different in the retest condition of sequence 2, compared to the test condition (test = 73.3810 Hz, retest = 65.6728 Hz).

2.6.6.4. Summary of test-retest analysis

The Test-Retest analysis shows that the word-retest condition is consistent with the results of sequence 1. The slope representing a categorization strategy defined by the word-test condition at the beginning of the sequence is maintained and not influenced by the sequence and highly correlated with the perceptual strategy used in word-retest condition. Midpoint values, despite the preceding 4th condition (humming) raises them to 70 Hz, appears not to be affected by it and they are kept constant and equal to the word-test condition.

Test-retest analysis pertaining the humming condition highlights that the categorization strategy at the beginning of the sequence 2 is not correlated with the one at the end of the sequence 2, after having categorized the word condition in 4th position. Slope values are not significantly different in test-retest phases, but the LMEMs' analysis has revealed that the slope in humming-test condition in sequence 1 is not in a linear relationship with slope values of the other conditions. This further supports the idea that the categorization of humming stimuli, based on a linguistic function, is fulfilled with different strategies compared to the categorization of stimuli with segmental information. This is, in final instance, again supported by midpoint analysis for the humming test-retest comparison highlighting a considerable rise in midpoint position in the retest condition, at about 70 Hz (middle frequency).

2.6.7. Correlated Measures of Individual Differences

During the experimental session two questionnaires: The Autism Spectrum Quotient (AQ) (Baron-Cohen et al., 2001) and the Interpersonal Reactivity Index (IRI) (Davis, 1983, Albiero et al.,2006) (see Chapter 1, paragraph 1.6.3 for a full description have been administered.
Because F0 conveys implicit and paralinguistic information together with the linguistic properties, it is conceivable that the presence of social impairments associated with the spectrum of autism correlates (e.g., scales: Ab_Soc, and Comunic) with difficulties in the categorization process of pitch. On the contrary, the ability to assume the interlocutor's perspective (e.g., PT scale) and higher empathic responses (e.g., EC scale) to other individuals' experiences should correlate with better performances in the categorization task.
The remaining two scales of the original AQ test, specific addressing the property to direct an extreme level of attention to details (Att_Dett) and to be resistant to disengagement (Sp_Att) can interestingly relate to the categorization process two-ways. As typical measures of the spectrum they can display a trend in the same direction of the two social scales (Ab_Soc, Comunic) and be associated with difficulties in the categorization task, or, on the contrary, given the fact that the task is autonomously performed at a computer and does not involve a real relational aspect of communication higher levels of attention directed toward one single parameter moving during the experiment, F0, can make the task easier for the participants. I predict therefore that these two subscales can correlate in both directions with the perceptual task, but negatively correlate with PT and EC scales.
Given the nature of the presented stimuli, the presence of the humming condition, and of the second sequence presenting stimuli from low quantity of segmental information to highly recognizable segmental information, I hypothesize that the subscales addressing imagination can play a role when a bottom-up categorization strategy has to be instantiated with stimuli that lack necessary segmental information. Sentence modality in the humming condition can require a certain level of the ability to imagine a statement or a question in association with descending and rising pitch.
The subscales addressing imagination in both tests (FS in IRI, Imag in AQ) should then negatively correlate between each other. In final instance, the subscale addressing the personal distress (PD) can be a general indicator, independent from traits belonging to the spectrum, of how distress during perceptual processing in communication can affect both top-down and bottom-up strategies when a stimulus has been identified as belonging to a category or when on the contrary on the basis of extrinsic reference a category has to be formed. Moreover, the literature on the IRI test highlights that PD negatively correlates with PT, meaning that individuals highly concentrated on their response to critical situation take less into consideration other individuals' positions and reactions. PD, in this way, should then negatively correlate with PT and EC, and positively correlate with Ab_Soc and Comunic.
The analysis aims to observe whether the high inter-individual variability displayed by subjects in sequence 1 in intercept position in slope and midpoint models can be explained by individual personality traits, pertaining to social communication.
The Principal Component Analysis (PCA) of the 9 subscales (5 belonging to the AQ test, and 4 to the IRI test) Is, thus, performed to test whether a correlation is present in each sequence, independently for the slope and for the midpoint, with the intercept's random effects of the most complex model fm3.
The PCA generates 9 orthogonal vectors expressing a combination of the 9 dimensions that, going from the 1st to the 9th, decrease in the proportion of explained variance in the data, calculated on the tests' scores of all 37 participants. I tested the correlation between the 9 PCA vectors and the random intercepts in the fm3 models, calculated only with 3 conditions carrying linguistic segmental information since the humming condition is not in a linear relationship with the other 3 conditions in the sequences. I start testing from the

PCA vectors explaining the highest portion of variance and shift downward (from PC1 to PC9) in the case the comparison is found non-significant.

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9
S.D.	1.5044	1.3579	1.1337	1.0321	0.9362	0.8134	0.6909	0.5849	0.4300
Prop of Variance	0.2515	0.2049	0.1428	0.1184	0.0974	0.0735	0.0530	0.0380	0.0205
Cumulative Prop	0.2515	0.4564	0.5992	0.7175	0.8149	0.8884	0.9415	0.9795	1.0000

Table 2.6.27 Principal Component Analysis of covariate measures

Table of Pearson's' Correlations											
Sequence 1						Sequence 2					
Random intercepts of **slope**-fm3 model						Random intercepts of **slope**-fm3 model					
Vector	t	r	df	p.value	C.I	Vector	t	r	df	p.value	C.I
PC5	1.6066	0.4070	13	0.1321	-0.1330 0.7607	PC8	-2.3748	-0.5654	12	0.0351	-0.8431 -0.0498
Random intercepts of **midpoint**-fm3 model						Random intercepts of **midpoint**-fm3 model					
Vector	t	r	df	p.value	C.I	Vector	t	r	df	p.value	C.I
PC1	-2.5176	0.5725	13	0.0257	-0.8388 -0.0852	PC8	3.0545	0.6614	12	0.0100	0.2015 0.8823

Table 2.6.28 Correlation of PCA dimensions with slope of the fm3 models

Loadings in the 3 emerged vectors					
PC1		PC5		PC8	
PT	-0.3728	PT	-0.5770	FS	-0.6100
Immag_30	-0.1518	Sp_Att_30	-0.3151	Immag_30	-0.3047
Att_Dett_30	-0.1468	PD	-0.2896	PT	-0.2108
Ab_Soc_30	0.1195	Immag_30	-0.2550	PD	-0.1552
EC	0.2271	EC	-0.0844	Comunic_30	-0.1068
Comunic_30	0.3192	Comunic_30	0.0084	Att_Dett_30	0.0120
Sp_Att_30	0.3945	Ab_Soc_30	0.1516	Ab_Soc_30	0.2264
FS	0.4705	FS	0.2238	Sp_Att_30	0.3240
PD	0.5213	Att_Dett_30	0.5819	EC	0.5466

Table 2.6.29 Composition of PCA dimensions that significantly correlate with fm3 models

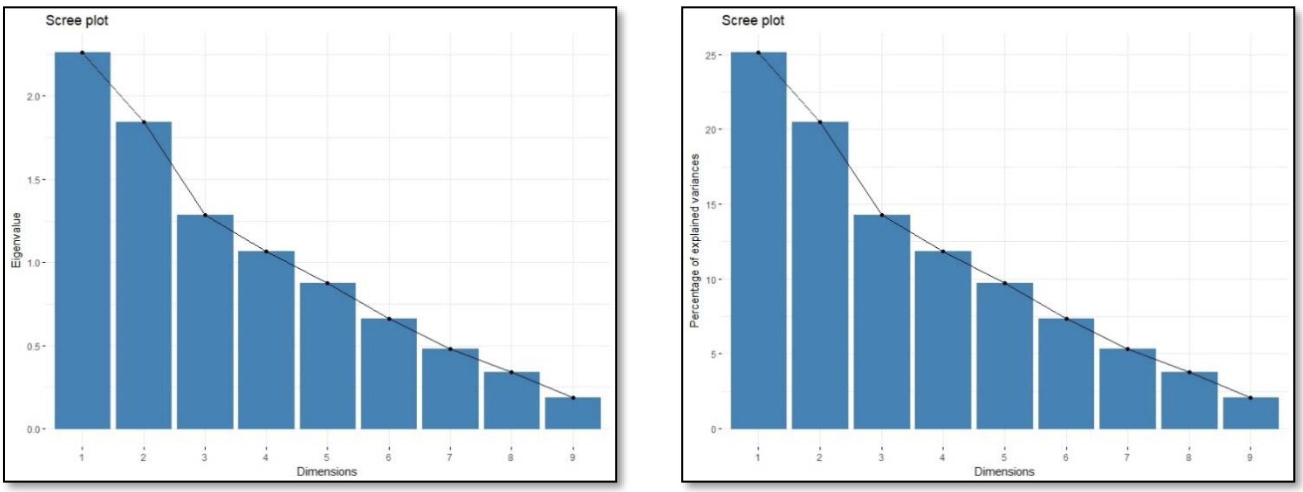


Figure 2.6.19 PCA dimensions' proportion of explained variance

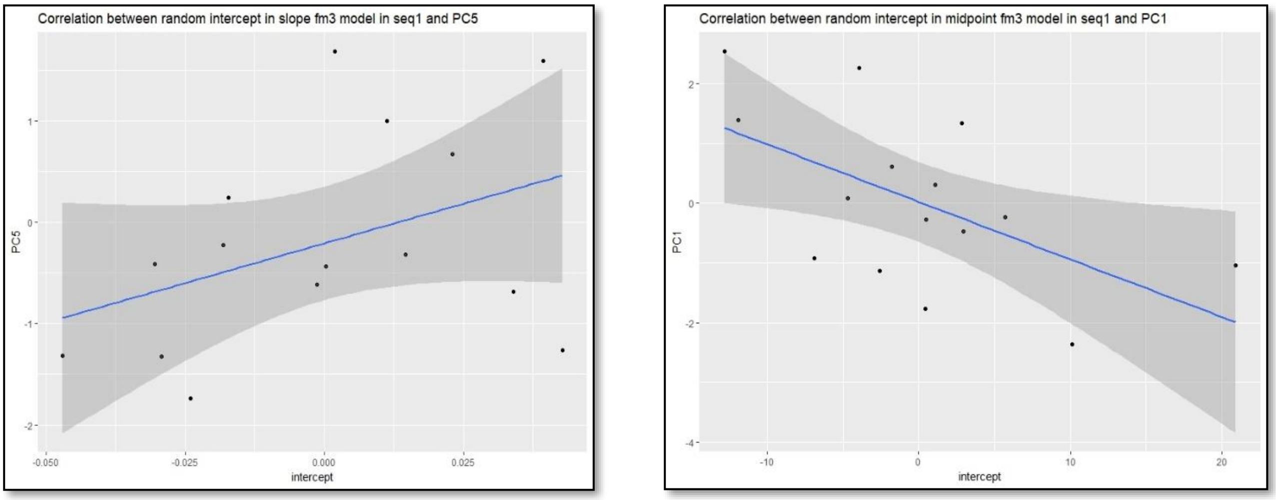


Figure 2.6.20 Significant correlations of PCA dimensions and fm3 model in sequence 1

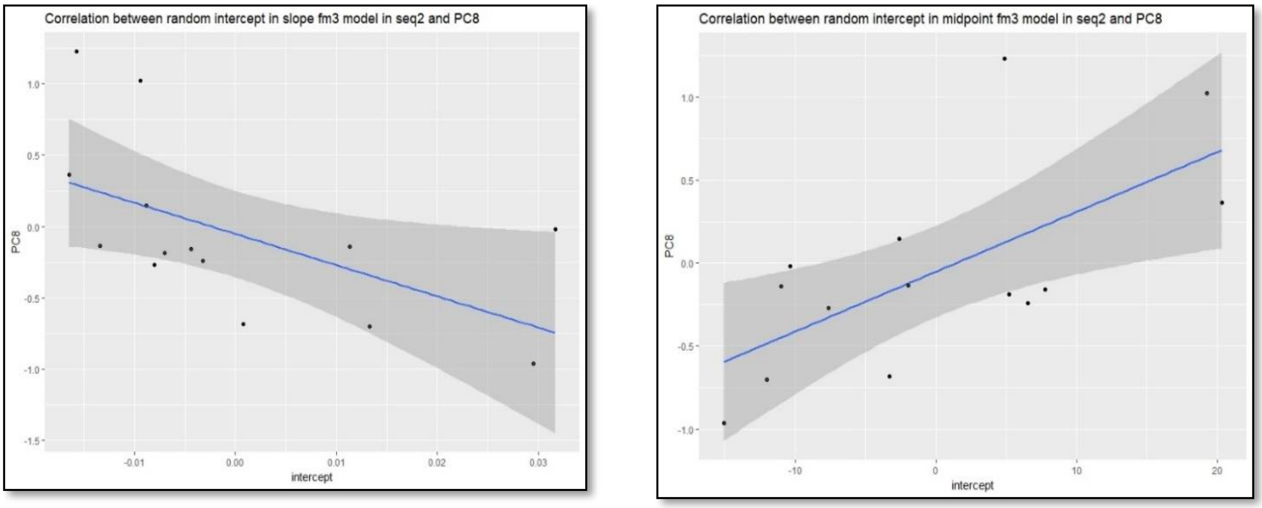


Figure 2.6.21 Significant correlations of PCA dimensions and fm3 model in sequence 2

2.6.7.1. Summary of PCA of correlated measures of individual variability

The aim of the PCA analysis was to observe whether the great variability in intercept position measured by random effects in LMeMs correlated with individual measures of easiness/impairments in the handling of social interactions adopting two tests: AQ (Baron-Cohen et al., 2001) and IRI (Davis, 1983; Albiero et al., 2006). I considered the PCA of the total number of 9 subscales present in the tests and the following outcome has emerged:

1) a non-significant tendency between the slope intercepts of the fm3 model in sequence 1 and PC5;

2) a strong significant correlation between the midpoint intercepts of the fm3 in sequence 1 and PC1;

3) a significant correlation of both slope and midpoint intercepts of fm3 models in sequence 2 and PC8;

The vector PC8 explains only 4% of the variance in relation to the 9 subscales and has an eigenvalue by far below the value of 1. This implicates that none of the subscales significantly explains the variance in the data; therefore, I do not discuss here possible relations between the order of the 9 dimensions within the vector in relation to the intercepts' values in slope and midpoint fm3 models. Nonetheless, it is interesting to highlight that the models in sequence 2 are the ones having significant fixed effects of the sequence on the values of slopes and midpoints and therefore it can be noticed that slope and midpoint values, representing a categorization process and the set of category boundaries, show to be homogeneously linked with the properties of the stimuli and to be less dependent from the individual variability of participants' social skills.

In sequence 1, results highlight a non-significant effect when testing the correlation of slope's intercepts and PC5, which explains the 10% of the variance and has an eigenvalue approximately reaching the value 1.

Despite the non-significant effect, that could be in function of the limited number of observations, the order of the 9 dimensions in the vector can be explainable on the basis of predictions, and therefore it can be further discussed.

Low values in the intercepts of the slope model, meaning that at the beginning of sequence 1 the perception of pitch movement is not clearly categorical, correlates, in order, with high values of the Perspective Taking (PT) scale, high values in the subscale measuring the resistance to Attentional Disengagement (Sp_Att) and high values of Personal Distress (PD). On the contrary, high attention to local details (Att_Dett), a high score in the Fantasy scale (FS) tendentially correlate with high values of intercepts, corresponding to high values of slope in the word condition of sequence 1.

The most relevant correlation is the significant correlation with the midpoint intercept in sequence 1 and the vector PC1, explaining 25% of the variance proportion in the tests and having an eigenvalue above 2. The correlation is negative, so participants, having high midpoint intercepts in sequences 1, are participants with high values of PT, difficulties in fantasizing (Immag) and paying extreme attention to local details (Att_Dett), on the contrary, participants with low values of midpoint intercepts in sequence 1 display high values of distress (PD), ability to fantasize (FS) and resistance to attention disengagement (Sp_Att). Interestingly the variables measuring relational skills play no relevant role in comparison with the previous one in relation to the midpoint positioning at the beginning of sequence 1.

If one assumes what previously discussed about the nature of the processing of the two sequences, meaning sequence 1 clearly activates a top-down linguistic processing on the basis of available representations of pitch categories in the native language, and on the contrary sequence 2 strongly activates a bottom-up processing of extrinsic reference across trials and instantiates a generalization process about pitch categories, it is possible to see that steep slope values correlate with attention deployment toward the properties that allow the identification of the categories, and a high ability to mentally recreate the dialogical function of pitch in determining sentence modality during perception (FS). It is important to remember that the task despite being centered on the linguistic function of pitch is performed alone by participants at the computer and it is highly repetitive. This interestingly emerges in the opposite way, observing participants having low values of intercepts, therefore not displaying a strong categorical perception in the first condition, that is clearly speech-like, having high values of PT the reflects a process of doubting on the original decision, once several trials are proposed in the categorization task, and the fact that perhaps the task is considered more for its communicative representation. High values in the Sp_Att scale correlating with low values of slope can highlight a difficulty in the reanalysis of each trial once a decision has been taken for the previous one in a limited range of variation of the stimuli and causing a similar rating of the trial as belonging to either one of the two categories with similar proportions, and reflecting indecision.

The subscale Att_Dett which appears to facilitate the decision-making process linked to slope values correlates with the shifting to higher values of the midpoint. It can be explained supposing that participants paying extreme attention to local details instantiate a process of extrinsic bottom-up reference even with the categories are fully recognizable in the word-condition. The process of comparison across trials then raises the position of the category boundary. If the subscale Sp_Att can be interpreted with difficulty in analyzing several trials continuously, then it is in accordance with the fact that participants that do not instantiate repetitive comparisons across trials set the category boundaries to a lower frequency-values.

In final instance, it is interesting to observe how high values of PD and FS and related to low values of the midpoint.

The positive correlation between these two scales is discussed in the literature pertaining the IRI validation and factorial analysis (Cliffordson, 2001, 2002), even though a negative correlation of PD with all the remaining scales has been found (Coke et al., 1978; Davis, 1980).

It is assumed that PD reflects the tendency of the participant to center the (communicative) environmental situation on him/herself, and his/her tendency into the non-consideration of other individuals' perspectives and therefore the enhancement of distress caused by the recognition that when the environmental situation becomes critical a solution has to be found autonomously and independently by the participant alone.

This can correlate with the tendency to speculate and fantasize with the possible outcomes.

In the specific task, the environmental situation that has to be solved is the categorization task itself, which is performed autonomously and independently by the participant alone that also has to mentally recreate a communicative dialogical situation where pitch is modulated to convey sentence modality. It is reasonable to interpret that higher values in personal distress can correlate with a "quicker" decision-making process about the category boundaries, with higher values in Sp_Att highlight the difficulties of the reanalysis of multiple stimuli with little variation and therefore leading to lower midpoint values.

2.7. Discussion

The hypotheses exposed in the introduction focused on the prediction that sequence 1 would instantiate a top-down identification (classical meaning) task in which the category of rising or descending pitch contours aligned in sentence/word-final position on the last syllable had to be recognized and associated to the respective sentence-modality function: statement vs. question. This process was assumed to be increasingly more difficult as the recognizability and access to meaning of segments decreased along the sequence.
The final condition “humming” was predicted to either linearly be the most difficult type of stimuli to be identified, or on the contrary to be so poor of linguistic properties to shift the processing to a non-linguistic one. In the second case, it would have been easier to identify based on acoustic properties, but its categorization would have shown significant differences with the linguistic trend.
The linguistic computation and its degree of difficulty are observed through the trend of slope and midpoint values of the fitted sigmoid curves in each categorization task; therefore in sequence 1, it was expected: a linear decrease of slope values and a linear increase of midpoint values, and a two-way hypothesis for the humming condition. The steepness of slope values is also associated with a predicted categoriality of perception, so the higher the slope values the similar the perception with the prediction of the classical CP paradigm.

The results in sequence 1 highlight, firstly that only one participant had to be removed as outlier for not correctly identifying the stimuli only in the humming condition in position 4. The results in sequence 1 further highlight no significant difference between slope and midpoint values in the test and retest phase of the word condition at the beginning and the end of the sequence. Overall this highlights that the way participants perform during the sequence is consistent and reliable.
Then, considering the sequence without the retest-phase and including the humming condition, no sequence effect is found for both slope and midpoint values.
There is no significant difference between slope values across conditions, and there is high variability between individuals in the intercept position of the model in condition 1, that does not correlate with the trend of the model. The high variability indicates that single participants have a personal linguistic criterion that does not change in function of segments along the sequence, but that is transferred to the other conditions along the sequence since values are not significantly different and highly intercorrelated.

Similarly, observing midpoint values, the sequence’s order does not predict a linear change, and there is interindividual variability in the intercept’s values, not correlated with the beta of the model. Moreover, the humming condition shows to have a significantly higher midpoint value compared to the word condition, setting on the half-frequency range available in the whole battery of stimuli, that is not correlated with the midpoint values of the word condition at the beginning of the sequence. Unexpectedly midpoint values in the foreign condition are significantly lower than the adjacent condition.
When removing the humming condition from the model, hypothesizing a different type of perceptual processing of pitch without segmental information, only the midpoint values show to be in a linear relationship, considering a random intercept position, and the sequence predictor appears to lower the midpoint values. This again appears to be guided by the foreign condition.

Considering only the models with 3 conditions (removing the 4th condition) the results further indicate that the high initial variability of slope and midpoint values in sequence 1 is related with individual differences in aspects of communication associated to the construct of Theory of Mind, such as empathy, the ability to

assume the interlocutor's perspective and the amount of distress perceived in critical situations. In particular, a stronger relationship is found for midpoint's intercepts values modulation. Here, the presence of the combination of the willingness to assume the interlocutor's perspective associated with a difficulty in imaging fictional facts correlates with high midpoint's intercept values. The combination of the easiness of imagination of fictional facts with the presence of high values of distress during critical and difficult situations correlates with low midpoints' intercepts values.
The connections between slope's intercepts values and social traits shows to be less strong (p = 0.13 and PC5 explaining the 10% of variance) but it indicates that the tendency to assume the interlocutor's perspective correlates with low values of slope and an atypical attention to local details, usually present in individuals with characteristics of the spectrum of autism, increases the slope values.

Overall these results indicate that the top-down linguistic activation has a strong impact on slope values' modulation and that participants, once they have identified the two available categories in the word-condition, can activate the same linguistic identification process even when segments are removed. The degree of categoriality of pitch perception expressed by the slope values show to be tightly bound to speech but also influenced by individual variability in communication skills.

The presented literature that has investigated CP of intonation is primarily concerned on the nature of pitch processing as fully belonging to speech processing, and/or how and to what extent. The results pertaining sequence 1 which shifts from an experimental paradigm analogous to the ones used before, to the use of vocal but not linguistic stimuli, highlights that the two categories emerge as speech categories expressing sentence modality and are identifiable. The degree with which the two categories are distinguishable and the position of the two categories in the participants' perceptual space is on average set by native language's structural properties, but it is highly variable between individuals, and this can be an explanation of the nature of "quasi-categorical" perception.
The presence of interindividual variability appears to be in line with Schneider's observation on the nature of pitch categories, as she defines them "half-closed" categories. This is visible in the data in the slight non-significant increase of slope variance in the "foreign" condition, and the significant lowering of the midpoint in this condition.

At group-level, the perceptual identification strategy of sentence modality is not affected by the degree of recognizability of segments, but the way the identification strategy is transferred to the conditions following the first one appears to vary across individuals. This can reflect the individual variability in processing the tune-text association and the pragmatic interpretation of pitch modulation, that results in the classical studies in "quasi-categorical" perception, and the unstable findings of magnets effects in some cases emerging only when identifying statements and in other cases emerging also in questions' identification.
The role of pitch range appears to affect the midpoint position in particular in the humming condition, that is not related to the linguistic conditions and shifts to the half range.

Looking only at the data of sequence 1 the midpoint values appears to be more affected by the experimental manipulation compared to the slopes as the Multistore Model predicts. In fact, removing the humming condition the sequence slightly instantiates an experience's effect not given by the available long-term representation of sentence modality but from the fact that the task is repeated along the conditions and therefore a short-term representation of sentence modality, guided by the experimental stimuli, also becomes available to the participants. This helps the identification process reducing the number of extrinsic references comparisons and therefore lowering the midpoint.

Thus, against the starting hypothesis that the sequence would become increasingly harder to be decoded, the results highlight that the speech strategy is top-down kept constant, and that a small, bottom-up effect experiment dependent affects the midpoint positioning.
Together with the fact that the processing of the humming appears to be different on the basis of midpoint positioning, it is possible to infer that slope and midpoint reflect two different cognitive processes. One process reflects the top-down linguistic guided activation and is related to slope values. The second one reflects a bottom-up evaluation of the multiple acoustic cues co-occurring with pitch modulation, comprehensive of pitch range and tune-text association and recognizability of text. The second cognitive process affects the midpoint positions, therefore the perceptual categories' boundaries. In addition, it appears that bottom-up perceptual strategies strongly correlate with individual pragmatic/social abilities.

In Sequence 2, pitch conveying sentence modality has to be categorized with stimuli varying from the humming-condition to the word-condition, where the variable called co-text (the predictor expseq2 in the analysis) varies from unrecognizable segmental information to fully recognizable segmental information.

The first observation is that the results diverge from the ones in sequence 1, and 4 participants are not able to categorize the humming stimuli on the basis of a linguistic criterion, and they had to be excluded. The second evident observation is that the modeling of slope and midpoint values within the sequence is not in a linear relationship with the linguistic conditions. This straightly points to the fact that pitch only, aligned to vocal spectral information, is not categorized with the same process adopted in the word condition in sequence 1, and shows differences with the humming condition appearing in position 4 in sequence 1, since in this case no long-term linguistic representation can be activated to identify the stimuli.

The participants who are able to solve the task and that can categorize the stimuli have to instantiate a bottom-up categorization process of the stimuli, guided by extrinsic references across trials. The generalization process that has begun in condition 1 has to be kept in short-term memory and re-used in order to proceed in the sequence and improve in the categorization task, as predicted by the Multistore Model.

Midpoint values in the humming condition reflect the midpoint values found in sequence 1, corresponding to the half-range of available frequencies in the experiment. Interestingly the midpoint values correlate only between the direct adjacent conditions. This further supports the emerging evidence that the midpoint setting reflects a bottom-up perceptual evaluation.
The slope values of the humming condition in 1st position in sequence 2 are not significantly different from the ones calculated in the 1st word-condition in sequence 1. However, they do not correlate with the following conditions, and more interestingly they are not correlated with the slope values of the retest phase, when only looking at test-retest data, possibly indicating that the top-down linguistic strategy once is activated within the sequence is adopted in the retest phase, but it is not available to the participants in condition 1.
Furtherly, slope values are interconnected across linguistic conditions, supporting the evidence that slope values reflect a top-down activation of linguistic representation in long-term memory.

The third relevant observation is the fact that once the humming condition is removed from LMEMs the sequence, and therefore the omni-variable "co-text" linearly predicts the top-down activation of linguistic information. In fact, slope values in the foreign condition in position 2 in sequence 2 are predicted by the factor expseq2 in the model, and random variation of the intercepts are related to random variation of beta in the model, indicating that the strategy adopted in this first unusual linguistic condition is the same adopted for native-language stimuli categorization only not fully properly functioning.

The change of processing is highlighted by the drop of slope values between the humming and foreign conditions in position 1 and 2, and from the linear increase up from the foreign condition.
The foreign condition in position 2 then reflects a situation where the system is aware that available linguistic representations can be adopted to classify the presented stimuli, but it also becomes aware that bottom-up information lacks relevant information to be fully processed as speech.
Segments are hardly recognizable, and meaning is not accessed.

The categorization then of three linguistic conditions, from mostly unrecognizable to fully recognizable, reflects a process of perceptual training, where slope values linearly increase, and midpoint values linearly decrease.
In sequence 2 therefore, there is full evidence for the application of the Multistore Model where both long-term linguistic representations are activated, the acoustic features of the stimuli are evaluated, and the result of the evaluation is kept active in short-term memory to support a bottom-up categorization process.
Results of sequence 2, moreover, support the position of Lotto & Holt demonstrating that categorical perception can be started by the bottom-up sensory evaluation of the stimuli and reflects a speech property when the signal matches the expected linguistic representation, but it can also be created starting with non-full speech stimuli.

These results are also providing relevant evidence to the debate on whether linguistic pitch is processed as speech and/or shows analogous perceptual properties compared to segments.

The neuroscientific evidence highlights both processing differences on the basis of acoustic features of the stimuli: such as fast frequency transitions for consonants, being left lateralized and longer frequency transitions such as the ones required by the suprasegmental features being right lateralized; and processing differences based on a functional approach guided by language experience.

The DSM of Hickock and Poeppel, in particular, predicts that speech and non-speech acoustical signals can share a baseline decoding processing system that can evaluate the common acoustic structure of the signal, in a similar way higher order processes can distinguish pseudo and real word but to grasp the common word-structure.
What grants a different type of processing to speech then is the categorial discrete nature of speech mental representations. For Hickock and Poeppel, thus, there is a difference between speech perception and speech recognition.
The reported studies, in fact, highlight that pitch modulated as lexical tones in tonal languages, such as Chinese and Thai, aligned to Chinese segmental information or on the contrary aligned to Thai or English segments, is differently processed, and related to different brain areas in native speakers of the specific tonal languages and non-native speakers.

The correlation with linguistic categoriality of pitch perception is found in Wang, 2004 in differences between native speakers of Chinese and native speakers of English perceiving pairs of Chinese words with Chinese tones superimposed on them, and English words with Chinese tones superimposed on them, and having to state whether the tones were the same or different. A main activation difference in function of the native language of the speakers is the correlation of speech recognition with the left Anterior Insula in native speakers of Chinese, where pitch acts as a distinctive feature and with the right Anterior Insula in native speakers of English, where pitch is a suprasegmental feature at sentence-level.
Moreover, Xu et al., 2006 find a bilateral activation of the Planum Temporale related to pitch perception but a left lateralization when pitch has a discrete phonological representation in perceivers' native language, as in Chinese.
Authors interestingly propose that the Planum Temporale functions as a hub between top-down and bottom-up processing; thus, it functions as a hub for incoming sensory evaluation of a stimulus, and amodal mental representation of a discrete category.
Finally, lateralization has been found in relation to affect prosody processing, compared to linguistic pitch modulation processing, in the studies of Wildgruber, where it appears that when the language system is engaged, the left inferior frontal gyrus is predominantly activated, and when affective prosody is processed, the activation is bilateral and strongly activates the right frontal opercular areas.
Interestingly, the behavioral result concerning the processing differences in the two sequences is strongly in line with neuroscientific evidence highlighting that the link between speech-categoriality and top-down activation of long-term memory representations.

Sequence 1 shows that when the identification process on the basis of language properties is correctly activated, the same identification strategy is transferred to the other conditions, even the humming one.
On the contrary, sequence 2 shows that a categorization process based on linguistic properties requires the evaluation of co-occurring segmental information, that most probably is related to the left lateralization of pitch processing with linguistic function.
The parallel bottom-up and top-down activation is in line with the position of Xu et al., supporting the fact a specific area, the left planum Temporale, connects information from both directions and matches them together within a linguistic system.
In this experiment, it appears that the humming condition in position 1 in sequence 2 impairs the matching process and then as information increases along the sequence the linguistic processing increases.
Future neurophysiological studies can, therefore, investigate whether the linear increase within linguistic condition in sequences 2 correlates with an increase of activation of left lateralized linguistic processing of pitch in the left IFG, left planum Temporale, left Anterior Insula.
On the other side, shallow slope values and higher midpoint values can indicate a linguistic categorization strategy is not fully activated and the bottom-up process can then correlate with the right lateralized hemispheric activation.

Individual variability in communication and social styles can then relate to pitch processing in the way single individuals can generalize linguistic strategies during pitch perception – speech perception, based on bilateral activation – and can link the emerging category formed during perception with top-down activation of discourse functions such in this case, of BT, to express sentence modality – speech recognition, expressing discreteness and categoriality.

2.8. Conclusion

The nature of pitch is interestingly very much suited to investigate differences between speech and non-speech perceptual processes. This study supports the position that given the nature of its signal (slow frequency transitions, carrying either linguistic, paralinguistic emotional, and musical information, the required tune-text association when carrying a linguistic or paralinguistic function) the analysis of pitch processing can disentangle where the acoustic processing of sound matches higher order linguistic discrete representations and how language experience shapes the mapping between sensory driven information and amodal stored representations.

In particular the study finds that pitch can be both linguistically and non-linguistically evaluated. The degree of recognizability of segments plays a role in the type of activated processing highlighting that full linguistic activation relies on a multiple evaluation of cues, not only on F0 movements. Moreover, the study highlights that once the linguistic processing is active it can be transferred to successfully identify stimuli displaying less co-occurring segmental information than expected. On the contrary, a processing based mainly on the acoustic evaluation can fail the formation of a discrete category, linked to sentence modality, and when it successfully occurs it requires the extrinsic evaluation of several trials.

The results support a paradigmatic shift from the imperative seek for Categorical Perception in relation to speech and pitch processing to the adoption of a cognitive perspective interested in disentangling the information that concurrently enhances the formation of discrete categories and interested in observing which processing phases are linked to the formation of categories.
The results support the concept of a perceptual system that parallelly evaluates the sensory driven information with the activation of a top-down linguistic representation and support the position that in cases where a linguistic representation is not matched it can be formed on the basis of sensory comparisons across multiple trials as the Multistore Model describes.
The two distinctive natures of the processing show to be behaviorally observable investigating slope and midpoint modulations of sigmoid curves generated with a CP paradigm, in particular the results indicate the slope is correlated with top-down activation and the midpoint setting is related to bottom-up sensory evaluation of the stimulus.

In final instance, the results indicate that the CP of BTs Is not a process homogeneously found, with respect to the number of available categories, the positioning of their boundaries, and the presence or not of strong prototypical exemplars of each categories because of the nature of Pitch itself and the high interindividual variability found when investigating how the co-occurring information is integrated and a final pragmatic category is accessed.
In particular bottom-up stimuli evaluation shows to correlate with the social and communicative traits connected the presence of Theory of Mind.
This further opens possibility to observe pitch processing in the investigation of atypical populations.

Chapter 3
Categorical Perception & Discrimination Threshold's estimation

3.1. Outline and Introduction

In this Chapter, I investigate whether the perception of statements and questions categories in the conditions lying at the opposite ends of the continuum representing linguistic intonation in the previous chapter – word and humming conditions – displays true categorical properties adopting the full paradigm composed of the identification and discrimination task.

Due to practical reasons concerning the length of the experiment, in Experiment 2, both the identification and discrimination tasks are performed in only two conditions only in one experimental order, corresponding to the Sequence 1 in the previous experiment, without the Test-Retest comparison. This results in a simple longitudinal test composed of two conditions, starting with the linguistic one – word – followed by the – humming – condition.
In the previous Chapter, the sequence resulted playing a determinant role in the perception strategy, namely: the early perception of the linguistic condition (word) allowed to associate the presence of the rising contour in the absence of segments (humming) with the pragmatic interpretation of a question. The word condition in 1st position allowed in fact to maintain a linguistic modality also in the conditions in which the recognizability of segments was lower, or consistently impaired as in the humming condition. In first instance in the previous experiment, the linguistic perceptual modality was signaled by the ability to correctly identify the categories (in sequence 2, the reversed order caused the exclusion of 4 participants because unable to categorize the humming condition) and in second instance, by a strong significant correlation of slope values across and along the conditions. On the other hand, the midpoint values significantly differed in the word and humming condition and did not correlate. In particular, in both sequences, the midpoint value appeared to be bound to the acoustical middle range of the frequency modulation, when the linguistic material was insufficient, and to be bound to the antecedent condition in sequence 2 reflecting a perceptual learning process.

Therefore, the first aim is to test whether the findings in the categorization task in Experiment 1 stably hold also in this experiment removing the two intermediate conditions, specifically observing whether the slope and midpoint values in the two conditions correlate and are significantly different.

The second aim is to investigate whether the presence of a sharp category boundary in the identification task correlates with the enhancement of the discrimination sensitivity in the second task and whether the presence or absence of segments enhances or not the ability to discriminate pitch within categories and at the categories' boundary. Since segments have been found to support a linguistic interpretation of pitch movements, the perception of pitch movements in word condition has a high probability to be quasi-categorical. On the other hand, it has been shown that the absence of segments correlates with an increase in the sharpness of the sigmoid curve. Thus, the predictions concerning the effect of the humming condition on perception modalities for the second time are twofold. If the linguistic modality is accessible and transferrable to the identification task with the humming condition, the degree of certainty with which participants recognize the linguistic categories, expressed by the slope parameter, could decrease, since the stimulus is less informative; in this case this would signal a continuous shift from the category statement to question. The discrimination rate, in this case, should not be affected by the identification task. On the other hand, if the humming condition triggers a perceptual strategy analogous the one triggered by the humming condition in 1st position in sequence 2 in the previous experiment, forcing the participants to find first an acoustical criterion and to associate it with the linguistic interpretation, then the results in sequence 2 of the previous experiment highlight in the humming condition, a steeper transition between categories, respect to the linguistic conditions. Paradoxically, this perception modality could enhance the presence of Categorical Perception (CP) and therefore decrease to a greater extent the discrimination sensitivity within category and leading to the emerging of a discrimination peak.

The third aim is to investigate whether individuals displaying a combination of communicative traits, such as the ones emerging from the application of the Principal Component Analysis (PCA) of the covariate measures of individual traits belonging to the ASC conditions, and to the construct of empathy, display enhanced discrimination sensitivity in discrimination task, or easiness in the categorization task compared to others with an opposite profile.

For these reasons, I introduce in this experimental paradigm an estimation of the minimal discrimination thresholds of the participants. In several experiments, it is debated whether the minimum step in the discrimination task is either too short or too high biasing the discrimination sensitivity. Moreover, by keeping the discrimination step equal across conditions, it would not be possible to estimate whether the discrimination sensitivity varies in function of a full linguistic processing of pitch, or not. By using an adaptive algorithm reducing the step between the stimuli presented to the participant it is possible to observe whether the discrimination rate substantially improves in the presence or absence of a fully linguistic stimulus.

In second instance, it is possible to address if finer perceivers, often denominated "pitch listeners" are individuals strongly relying only on the acoustical properties of the signal, or if, on the other hand, are more efficient users of the linguistic categories, and whether this correlates with a very low presence of the traits known to characterize individuals with Autism Spectrum Conditions (ASC).

The final aim is to investigate whether the frequency range, spanning over about 140 Hz, plays a significant role in discrimination capabilities. Namely whether to the presence of anatomical and physiological constraints, such as the presence of the Critical Bandwidth (CB) represents a universal bias towards a more efficient discrimination of variation in statements compared to questions. This would add relevant information to the ongoing debate on the presence and amount of prosodic categories expressed through the modulation of the final rise, that in an increasing number of studies reports a more identifiable linguistic function of statements compared to questions. The presence of a perceptual magnet within the category statement (Schneider, 2012) could be related to the presence of more performing discrimination skills at a lower range of frequencies compared to a higher one. This type of information is not possible to obtain adopting the classical paradigm.

3.2. Method

3.2.1. Perceivers

All participants were informed before the experimental session about the overall procedure, the questionnaires they will have had to fill in, and their right of leaving at any time the experimental session, if uncomfortable. All participants signed written consent of voluntary participation before participating in the study. No form of monetary reimbursement was given to any participant, in the case they were undergrad students of the faculty of Cognitive Sciences, the overall participation's period of time was counted and recognized as experimental session time and could, therefore, be summed up and converted in useful credits in their study plan.

20 participants took part in Experiment 2. None of the participants of Experiment 1 took part in the second experiment. The participants were all women, native speakers of Italian, of age between 18 and 30 years old. They have all been raised as monolingual speakers, in northern-east of Italy. 12 perceivers of 20 were coming from Veneto, 2 from Lombardia, 1 from Friuli-Venezia Giulia, and 5 from the province of Trento (Italian speaking province of Region of Trentino-Alto Adige in Northern Italy). No one has reported hearing deficits.
The participants had to fill in before being physically called in the lab, an online screening questionnaire. As done in Experiment 1, I have selected participants not having any particular musical education nor a specifically high linguistic education, resulting in advanced proficiency levels in more than one language. The sample is intended to represent a population of individuals without any particular relevant training for F0 discrimination abilities. Since English as a second language is mandatory in the Italian education system, participants with knowledge of English as L2 have been allowed to the experiment.

Adopting the same procedure of Experiment 1, participants self-declared their level of proficiency in foreign languages according to the Common European Framework of Reference for Languages (CEFR) in 5 language-aspects: listening, reading, interaction, oral production, and written production competences. For each aspect, participants were asked to state their level ranging from the labels: "without any competence," A1, A2, B1, B2, C1, C2 language level. The total amount of participants responses has been added, and the average number of participants stating the same competence level in each aspect of each language has been calculated. The total number of answers has been averaged per language and converted in percentage for comparison.

3.2.1.1. <u>Participants of Experiment 2</u>

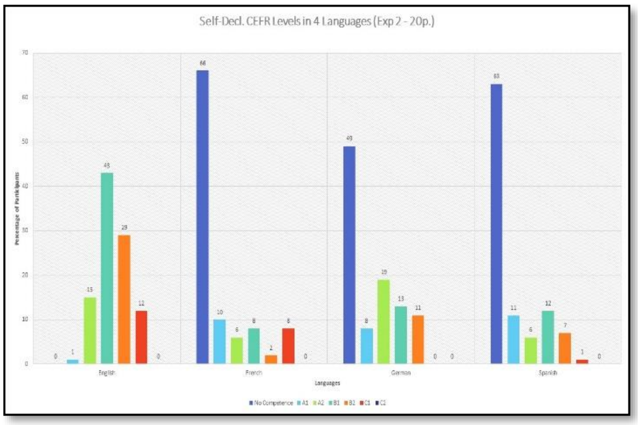


Figure 3.2.1 Distribution of CEFR's level in known L2

For what concerns the use of local dialect:

- 3 participants (15%) reported a minimal linguistic competence in their regional dialect in comprehension, and not to use it at all in production.
- 8 participants (40%) reported not to actively speak their regional dialect but to be extensively able to understand it.
- 7 participants (35%) reported productively using dialect only in a friendly and well-known environment, being able to produce simple sentences.
- 2 participants (10%) declared a broad use of dialect, declaring a comparable linguistic competence to Italian.

None of the participants underwent a prolonged and professional musical training.

3.2.2. Procedure

The experiment was run in a laboratory of the Department of Cognitive Science, of the Faculty of Trento, in the town of Rovereto in Northern Italy. Participants were seated in individual sessions in front of a computer in a quiet room at the presence of a research assistant. Each participant listened to the audio files through a pair of Sennheiser hd230 headphones and with the support of an external Sound Card: Echo Audio Fire 2 - Firewire 400. The procedure was launched on the MATLAB platform (2014b), with the aid of two specific toolboxes: Psychtoolbox (3.0.12.426025497) and Palamedes. The participant entered her responses on a keyboard. She was instructed to sit comfortably and pay as much as possible attention to the stimuli. The experiment was run in two sessions in the same day: firstly, the categorization and discrimination task were performed in the word condition, and in the second session the same procedure was adopted in the humming condition.
Between the two sessions, participants could take a long pause and shortly leave the lab and interact with the experimenter. Within the long pause, participants were asked to compile the two questionnaires – Interpersonal Reactivity Index and Autism Spectrum Quotient.

3.2.2.1. Categorization Task

At the beginning of the experimental session, participants were instructed on their task. The task was the same in the two conditions: they had to state after each stimulus whether they perceived the stimulus as having an intonation of an affirmative utterance or of an interrogative utterance. 10 trials of practice were played at the very beginning of each experimental session, with trials belonging to the respective condition.

3.2.2.2. Discrimination Task

The discrimination task was always performed after the categorization task of the same condition. The procedure estimating the discrimination threshold was run in each condition 3 times, with 3 different batteries of stimuli. All participants performed the discrimination task in 2 points of the perceptual space, corresponding to a set of stimuli belonging to the statement category and a set of stimuli belonging to the question category. For all participants out of the 1000 created stimuli, the procedure was run starting from the stimulus 100 (statement category), corresponding to a linear final rise of 14 Hz up from the first statement stimulus. Analogously, the starting point for the estimation procedure within the question category was set for all participant at the stimulus 642 corresponding to a linear rise of about 90 Hz from the stimulus 0.

The procedure spanned from the starting point within a range of 357 stimuli corresponding to about 50 Hz. The third starting point corresponded to the actual category boundary estimated in the previous categorization task. This point, therefore, varied across participants, but for all participants represented the true boundary in which they switched from clearly perceiving statements to clearly perceive questions.

The procedure randomly proposed in each block a sequence of 25 trials composed of triplets of the form AAB, ABA, or BAA, spanning at the first trial over the whole range of 50 Hz, and corresponding to 3 different keys on the keyboard. The stimulus A was always the starting stimulus of the procedure, and the stimulus B varied following the adaptive procedure. Thus, the stimulus A had always a lower boundary compared to the stimulus B, and the adaptive procedure moved backward from the maximum range to values always closer to the selected stimulus A. For each starting point the estimate procedure was run 5 times. Overall, each condition consisted of 15 blocks of 25 trials. Between each block, the participant could stop and take a short break. The participant started autonomously again the upcoming block.

Before the experiment in each condition, every participant could familiarize with the task in two blocks composed by 10 trials each, requiring to discriminate stimuli respectively within the range of the statement category and within the range corresponding to the question category. The try-blocks in this way spanned over the comprehensive range of frequencies allowing the participants to familiarize with the overall spectrum of frequency values adopted in the experiment.

Participants were required to identify within the triplet presented at each trial the different stimulus (B) among the 3. The adaptive algorithm on the basis of the correct response of the participant at each trial would then shift to a range lower than 50 Hz, diminishing, thus, the difference between the starting stimulus (A) and the different one (B). In case of a wrong answer, the algorithm represented a trial with the same selected stimuli.

At the end of each block, at the 25^{th} trial, the estimated minimum threshold value scored by the participant was saved. While perceiving stimuli, three white stars appeared on the screen, but participants were not obligated to stare at it without moving.

3.3. Results

20 female participants participated both to the categorization and discrimination task.

I firstly analyze as separate experiments the Categorization and Discrimination tasks to observe whether aggregated data support the recognition of two intonational categories in both word and humming conditions, and independently from the categorization task, whether the discrimination thresholds significantly decrease when the task is performed with stimuli corresponding to the midpoint areas of each participant. I, then analyze the relation between individuals' categorization parameters and discrimination's thresholds in order to evaluate whether participants showing an abrupter shift, from the perception of the category "statement" to the perception of the category "question", display also a significant increase of discrimination capabilities in the midpoint area, and decrease within categories' boundaries.

3.3.1. Categorization Task – Results

3.3.1.1. Outliers' removal

The categorization task has been performed in word and humming conditions by a total number of 20 participants. The curve fitting procedure on the results of each task in each condition generated a total number of 40 logistic curves, described by two parameters: slope and midpoints.
A first visual inspection of each of the fitted psychophysical curves, emerging from the categorization task in word and humming conditions, shows that no participant performed the categorization task inverting the categories (e.g. stimuli presenting a final rising contour identified as statements), or display extremely low values of slope signaling the non-recognition of questions and statements.
As in the previous chapter, the Log-Likelihood values (LL) of the fitted curves is adopted as indicator for the presence of single fittings that could possibly be identified as outliers.
A first observation of the distribution of LL values, calculated for each curve in both word and humming conditions, shows that there are two LL values distant from the mean and median values, approximating the value of -100, as reported in figure 3.3.1 and table 3.3.1. The same procedure for outliers' exclusion adopted in Chapter 2, is adopted here, therefore the data of participants fitting curve with LL values outside the Tukey limits are excluded.

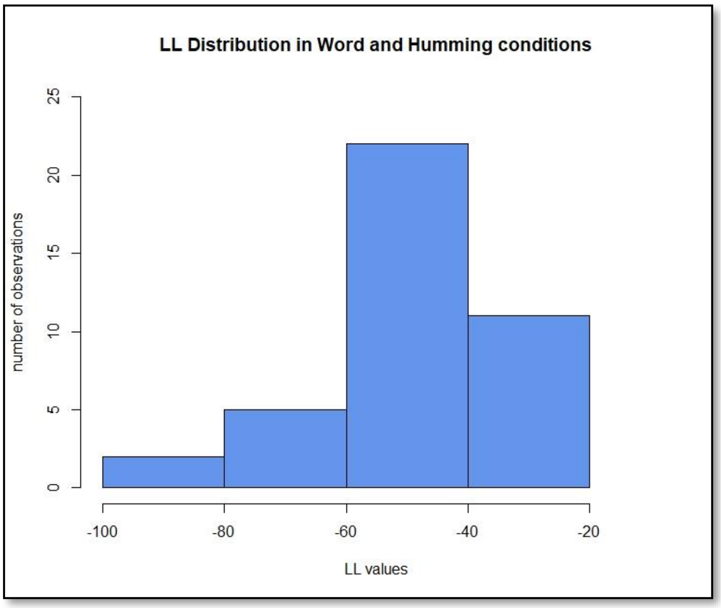


Figure 3.3.1Distribution of LL values across conditions

Overall Descriptive Stats of LL					
Min.value	1st Quantile	Median	Mean	3rd Quantile	Max.value
-94.07	-56.35	-49.91	-50.61	-39.21	-28.26

Table 3.3.1 Descriptive Statistics of LL values across conditions

Criterion for participants' exclusion – Outliers below the lower Tukey limit		
Higher Limit	Q3 + 1.5* IQR	-13.5
Lower Limit	Q1 – 1.5* IQR	-82.06

Table 3.3.2 Calculation of the lower Tukey limit - criterion for outliers' removal

There are two data points corresponding to the LL values of curves generated in the categorization task by nsub3 and nsub11, respectively in humming and word condition. The data of these two participants are excluded from further analysis of the categorization task, annotating that the fitted logistic curves displayed significant lower values of slope compared the average values of the sample.
The resulting dataset adopted in the categorization task's analysis is therefore composed by 18 participants, whose fitted categorization's logistic curves displayed values between -80 and -20.

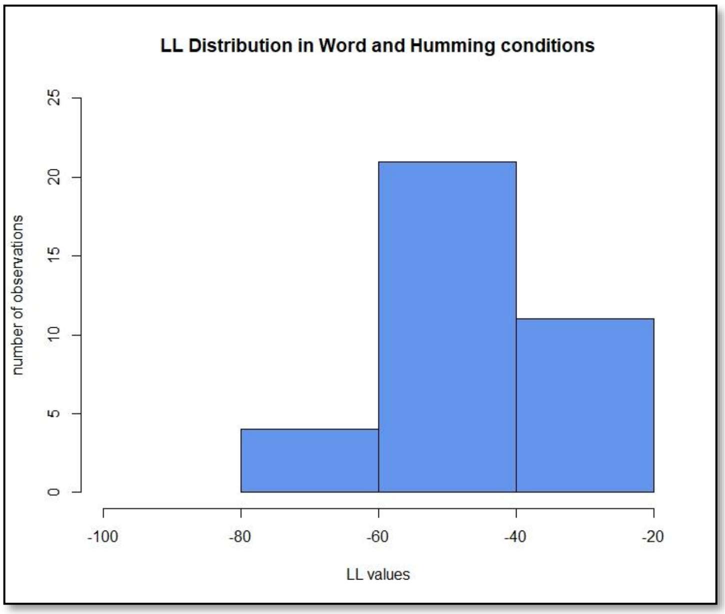


Figure 3.3.2 LL distribution after outliers' removal

3.3.1.2. Descriptive Statistics

Slope in Word and Humming Conditions							
Order	Conditions	Mean	S.D.	Var.	Min	Max	Obs.
1	word	0.0972	0.0264	0.0007	0.062	0.149	18
2	humming	0.1111	0.0271	0.0007	0.059	0.166	18

Table 3.3.3 Table 3.3.3Descriptive Statistics of slope values

Midpoint in Word and Humming Conditions							
Order	Conditions	Mean	S.D.	Var.	Min	Max	Obs.
1	word	61.8234	12.2336	149.6598	40.5990	70.8060	18
2	humming	68.5161	11.9982	143.9562	48.9750	87.4920	18

Table 3.3.4 Descriptive statistics of midpoint values

3.3.1.3. Correlation of slope and midpoint values in word and humming conditions

2.8.4.1.

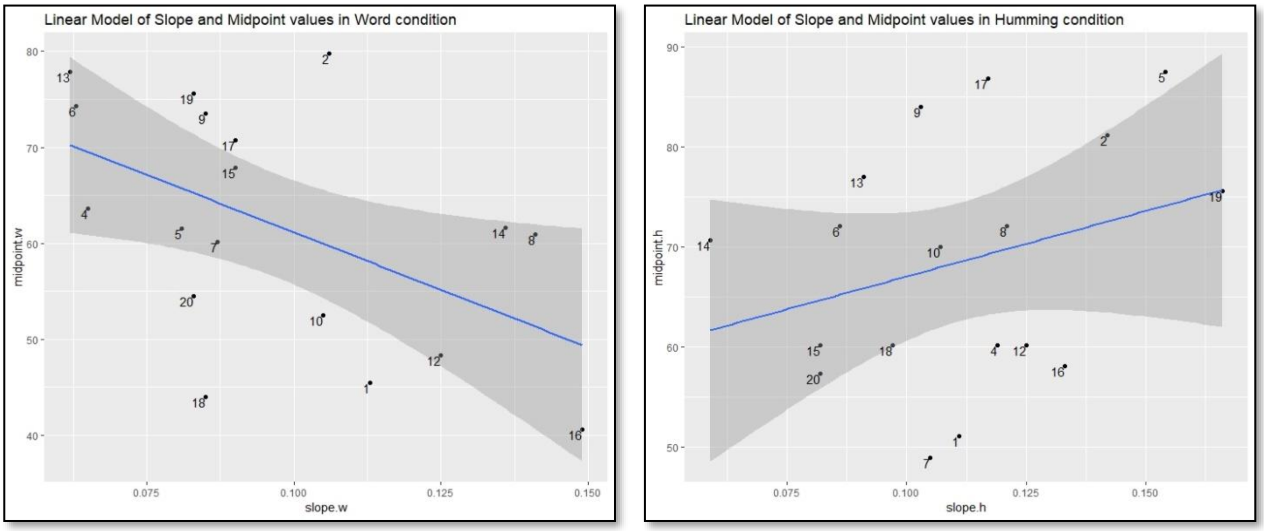


Figure 3.3.3 LM of slope and midpoint relation within category

Pearson's product-moment correlations between Slope and Midpoint						
Condition		t	df	Corr.value	p.value	Confidence Interval
1	Word	-2.422	16	0.5180	0.0277	-0.7931 -0.0674
2	Humming	1.2353	16	0.2951	0.2346	-0.1992 0.6697

Table 3.3.5 Correlation between slope and midpoint within category

3.3.1.4. Analysis of slope and midpoint values across conditions

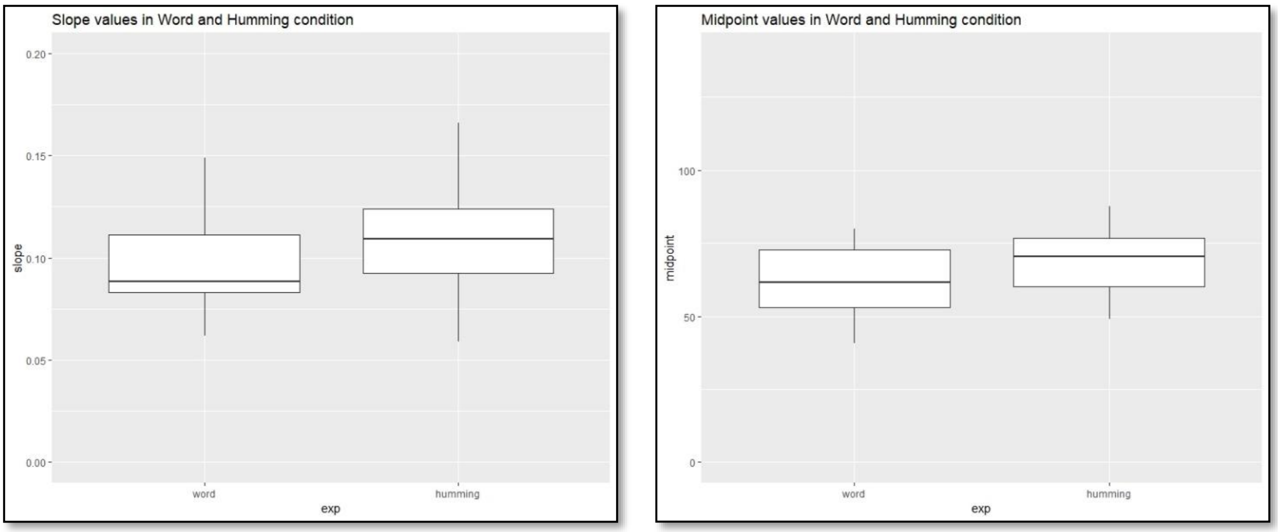


Figure 3.3.4Slope and midpoint values in word and humming conditions

The results of the paired T-Tests, reported in Table 3.3.6, highlight that the values of slope do not significantly vary in the perception of the word and humming condition, while the position of the midpoint significantly raises toward the half-frequency range as predicted. The data replicate the results of Experiment 1.

Paired T-Tests between Word and Humming conditions				
Slope	Word - Humming			
	t	p.value	df	C.I.
	-1.6269	0.1221	17	-0.0320 0.0041
Midpoint	Word - Humming			
	t	p.value	df	C.I.
	-2.8343	0.0115	17	-11.6744 -1.7108

Table 3.3.6 T-Tests of slope and midpoint values in word and humming conditions

The results of the correlation analysis highlight that the Midpoint value as expected is set on the basis of two completely different non-correlated criteria in the word and humming condition. On the other hand, the results of the Pearson's product-moment correlation test of slope values highlight the presence of a non-significant tendency. Therefore, the effect emerged in Sequence 1 in Experiment 1, signaling a strong significant correlation of slope values along the sequence is not fully replicated. The possible reasons are discussed in the discussion session below, nonetheless it is encouraging that a tendency of replication is present. (see Table 3.3.7)

Pearson's product-moment correlations between Word and Humming conditions						
Value		t	df	Corr.value	p.value	Confidence Interval
1	slope	0.3041	16	0.0758	0.7650	-0.4054 0.5241
2	midpoint	3.4983	16	0.6583	0.0030	0.2764 0.8607

Table 3.3.7 Correlations of slope and midpoint values across conditions

3.3.2. Linear Mixed-effects Models (LMEMs) with 2 conditions

In this section I adopt the same LMEMs analysis adopted in Experiment 1, investigating whether slope and midpoint values linearly increase or decrease across the two conditions, signaling a perceptual learning process. The models' testing follows the analogous hierarchical structure adopted in Experiment 1 and reported here below:

- fm3, the model comprehending fixed and correlated random effects (~ expseq + (expseq|nsub));
- fm2, the model comprehending fixed and where random effects are left free not to correlate (~ expseq + (1|nsub) + (0 + expseq|nsub)):
- fm1, the model with only the fixed effects and random intercept (~ expseq + (1|nsub)).

I consider then the cases where the fixed effect of the sequence predictor does not significantly explain variance in the data, and then only the models where random effects of the participants are modulated. I call, then:

- m3, the model considering only correlated random intercept and slope of the linear model (~ (expseq|nsub));
- m2, the model that considers the case where random effects are not correlated (~ (1|nsub) + (0 + expseq|nsub));
- m0, the basic model comprehending only the random intercepts (~ (1|nsub)).

3.3.2.1. <u>LMEMs of slope and midpoint between the Word and Humming conditions</u>

3.3.2.1.1. Slope

Model	df	AIC	BIC	Chisq	p.value
Random intercept (m0)	3	-152.13	-147.38		
Fixed Effect and random intercept (fm1)	4	-152.74	-146.40	-146.40	0.1065

Table 3.3.8 LMEMs of slope - model selecting (m0)

The ANOVA model comparison of slope values between Word and Humming conditions does not display any significant effect, in any model comparison, and m0 is the best model describing the data (beta0 = 0.104139 , SE = 0.004624, t = 22.52). Slope values are not significantly different across conditions and no variation of the intercept is related to beta values along the conditions.

3.3.3. Correlated Measures of Individual Differences

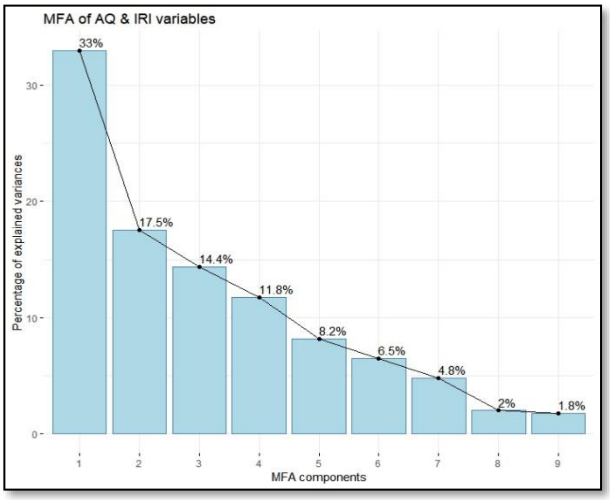


Figure 3.3.5 PCA (= MFA) dimensions' proportion of explained variance

	Dim1	Dim2	Dim3	Dim4	Dim5	Dim6	Dim7	Dim8	Dim9
S.D.	1.7194	1.2622	1.1353	1.0109	0.8681	0.7782	0.6629	0.4192	0.4068
Prop of Variance	0.3285	0.1770	0.1432	0.1135	0.0837	0.0673	0.0488	0.0195	0.0183
Cumulative Prop	0.3285	0.5055	0.6487	0.7622	0.8460	0.9133	0.9621	0.9816	1.0000

Table 3.3.9 PCA dimensions

The linear modelling of covariates PCA dimensions with participants' random intercept values within LMEMs models of slope and midpoint values across conditions, (estimated with the fm1 model) highlights that there is no significant correlation effect with any PCA components. The results indicate then that the main guiding phenomenon in determining slope and midpoint values within the word condition is highly linguistic, and less depend from the socio-communicative profile of a speaker.

3.3.3.1.1. Midpoint

ANOVA model comparison of midpoint values indicates fm1 as best model (beta0 = 61.823, SE = 2.856, t = 21.648; beta = 6.693, SE = 2.361, t = 2.834)

Model	df	AIC	BIC	Chisq	p.value
Random intercept (m0)	3	282.46	287.21		
Fixed Effect and random intercept (fm1)	4	277.49	283.83	6.966	0.008

Table 3.3.10 LMEMs of slope - model selecting (fm1)

The results of the LMEMs analysis of slope and midpoint values across conditions in the categorization tasks replicate the absence of a linear relationship between the modelling of slope values in the word condition and humming, in a sequence presenting the linguistic condition first. Participants shows a high variability in their degree of certainty of identification of the statement and question categories and the criterion adopted for the identification in word condition does not predict the criterion that will be further on adopted in the categorization of the humming condition.

In second instance, the modelling of the midpoint value, across two conditions only, one being the humming, replicates the results of Experiment 1. Despite the relationship the positioning of the category boundary on the basis of a presumed linguistic criterion does not correlate with the positioning of the boundary in the humming condition, which shows the usual tendency toward the half-range value, the midpoint values in word condition, randomly assigned within the group, can predict the behavior of the participants to shift it towards higher values when extremely low, and to decrease it toward the mid-range when extremely high in the word condition. This confirm that the position of midpoint is highly influenced by the condition of stimuli perceived in a prior task.

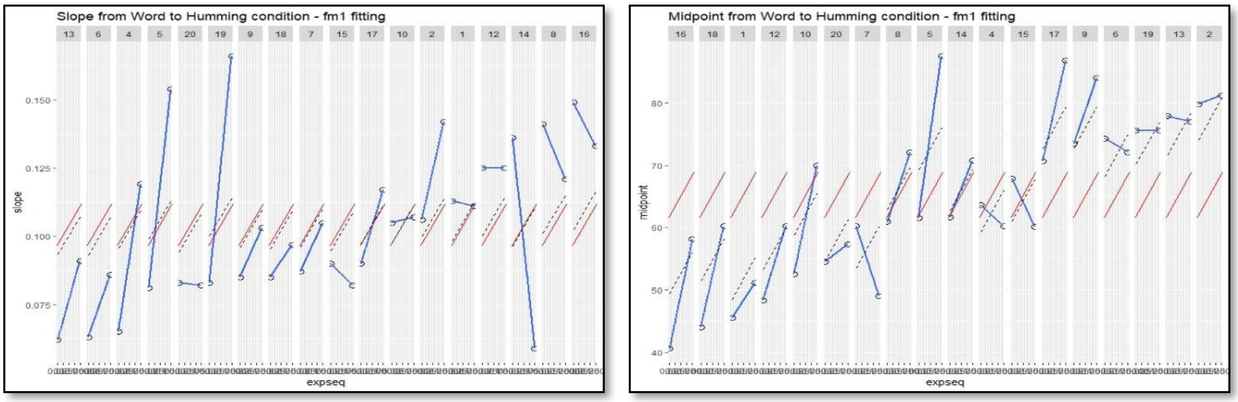


Figure 3.3.6 Longitudinal trend across conditions of slope and midpoint values

3.3.4. Discrimination task - Results

3.3.4.1. <u>Outliers' removal</u>

Out of the expected overall 600 observations, 545 successful threshold estimates have been collected considering the entire dataset, aggregating word and humming condition. The estimate procedure has been run in three "points" of the perceptual space of the participants, corresponding to 3 batteries of stimuli, diverging in average final rise f0 values and corresponding to 2 abstract categories (A) and (Q) and to the category boundary (M) (in the analysis below I use interchangeably the term "category" or "point" – referring to the position on the sigmoid curve to which the labels (A) (M) and (Q) correspond, also for data collected at the category boundary (M)):

1) Categorization of Affirmative Sentence (A)
   Level 100 = 14 Hz (F0 rise from the recorded Affirmative stimulus – Level 0)
2) Categorization of Interrogative Sentence (Q)
   Level 642 = 89 Hz (F0 rise from the recorded Affirmative stimulus – Level 0)
3) Transitional area between the two Categories (M)
   Estimated Midpoint of each participant.

The overall Threshold distribution is plotted in Figure 3.3.7

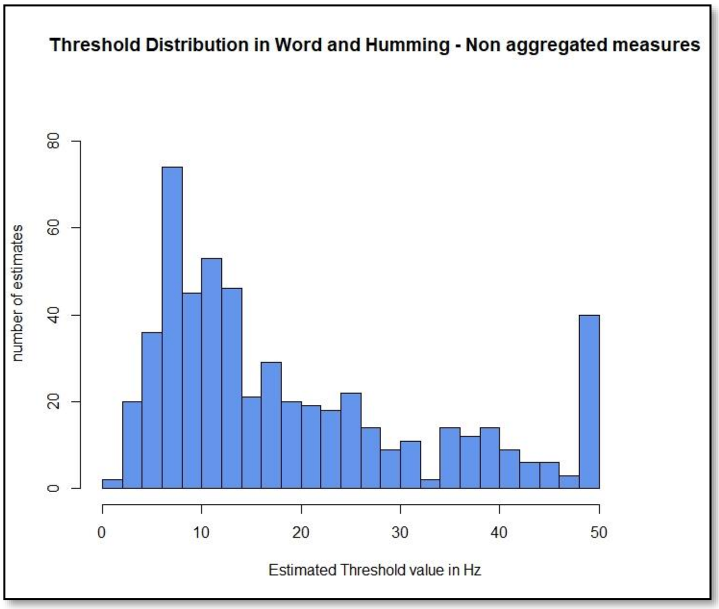


Figure 3.3.7 Overall threshold estimates' distribution

3.3.4.2. Descriptive Statistics of after the removal of ceiling data

Descriptive Stats of Threshold Estimates					
Min.value	1st Quantile	Median	Mean	3rd Quantile	Max.value
1.12	7.98	12.88	16.89	23.66	48.16

Table 3.3.11 Descriptive statistics after first data cleaning

Criterion for participants' exclusion – Outliers above and below Tukey limits		
Higher Limit	Q3 + 1.5* IQR	47.18
Lower Limit	Q1 – 1.5* IQR	-15.54

Table 3.3.12 Calculation of the Tukey limits

There are 5 observations corresponding to the max.value = 48.16 that could be classified as outliers with the same criterion adopted above. Since the raw data will be in the following steps of the analysis averaged per level and per participant, I choose to preliminary keep these 5 observations since it is not possible to assume with certainty they represent an error in the estimation procedure.

A first look at the Threshold estimates aggregated considering the whole dataset, and therefore not assuming any possible difference between the Word and Humming dataset, reveals that the threshold values estimated at the point (M), the personal category boundary of each participant in each condition, is not significantly lower than the values estimated within the two categories (A) and (Q) as the CP paradigm would predict.
Cleaned raw data are aggregated in Figure 3.3.9 below:

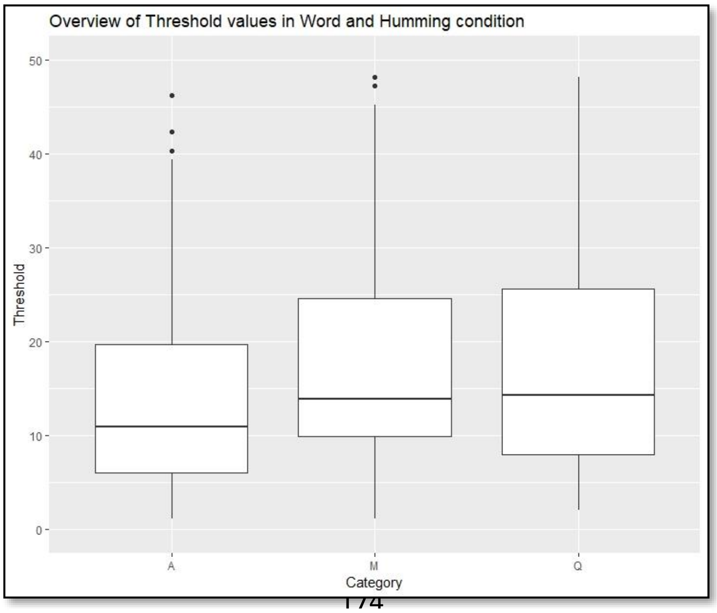


Figure 3.3.8 Overall threshold values in A M and Q, independently from condition - not averaged data

3.3.4.3. Descriptive Statistics of raw data

Descriptive Stats of Threshold Estimates					
Min.value	1st Quantile	Median	Mean	3rd Quantile	Max.value
1.12	8.96	13.86	19.19	26.60	49.14

From the distribution plotted above it is possible to observe that there is a consistent proportion of estimates that corresponds to the ceiling value of 49.48, or a slightly lower value. By visual inspecting the data, there are 34 observations reporting the ceiling value of 49.14 Hz. This finding would suggest that 6% of the times the participants were not able to discriminate a pitch difference of about 50 Hz. Given the fact that the estimated value in these cases is always the same and does not even vary in any value in decimal position, I consider these measures, cases where the estimate procedure was not correctly performed by the algorithm.
Therefore, I exclude them as outliers from the dataset.

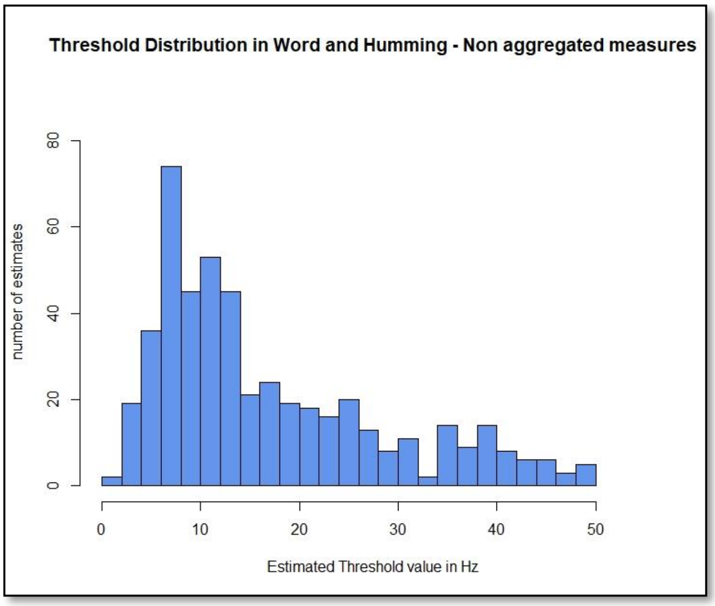


Figure 3.3.9 Distribution of threshold estimate values after outliers' removal

The resulting dataset is composed of 511 measures.
For one participant (nsub3), no threshold estimate is available assuming a starting point corresponding to the categorization of interrogative sentences (Level Q). Therefore, I also completely exclude this participant from further analysis. The dataset is finally composed of 491 threshold's estimates

3.3.4.4. Descriptive Statistics of threshold values in A M and Q points independently from condition

Mean values of Threshold Estimates in A M Q categories (all dataset)						
Category	Mean	S.D.	Var.	Min	Max	Obs.
A	14.62034	10.61530	112.6847	1.12	46.20	174
M	17.72589	11.02376	121.5232	1.12	48.16	163
Q	18.55636	12.66051	160.2886	2.10	48.16	154

Table 3.3.13 Descriptive stats of threshold estimates in A M e Q points

Results of (type 3) ANOVA between the categories indicate that the minimum estimated thresholds are sensible to a main Category effect F(2,488) = 5.51; p = 0.004. The Levene's Test for Homogeneity of Variance` is not significant. Post-hoc t-tests between levels highlights that the threshold-values estimated within the (A) category are significantly lower than the estimated threshold values within the other two categories: A vs. Q: t = -3.039; p = 0.008; C.I. (-6.49 - -1.38); A vs. M: t = -2.63; p = 0.013; C.I. (-5.43 - -0.79).

FDR adjusted post-hoc Kolmogorov-Smirnov (ks) tests between categories highlight that, the threshold values at the (A) category are significantly lower than within the other 2 categories: A vs. Q: ks = 0.18; p = 0.0155; A vs. M: ks = 0.22; p = 0.0009.

In the next section, the dataset is subdivided considering the observations collected in the Word and in the Humming condition. The word condition dataset consists now of 230 observations and the humming condition dataset of 261.

3.3.4.5. Descriptive Statistics of threshold values in A M and Q points in word condition

Mean values of Threshold Estimates in A M Q categories only in word condition						
Category	Mean	S.D.	Var.	Min	Max	Obs.
A	13.25438	9.546722	91.1399	3.08	42.28	89
M	20.81662	11.668827	136.1615	2.10	48.16	71
Q	22.44200	13.596115	184.8543	2.10	48.16	70

Table 3.3.14 Descriptive stats of threshold values in word condition - not averaged

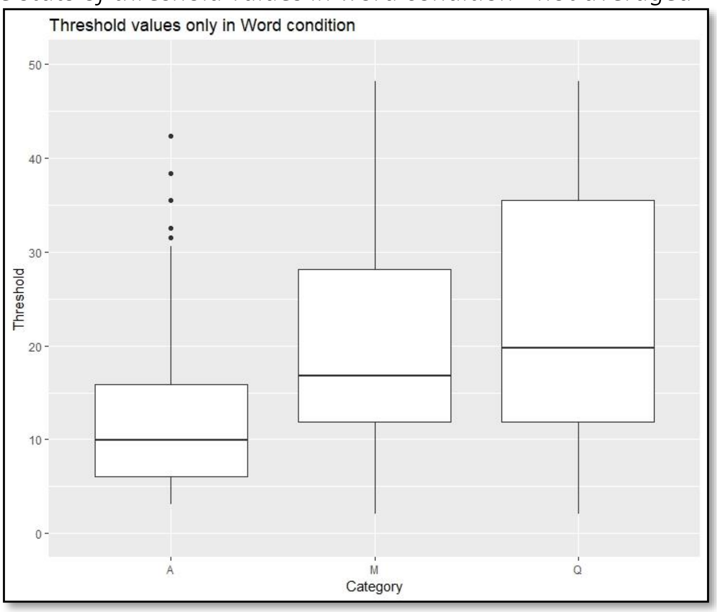


Figure 3.3.10 Threshold values in A M and Q points in word condition - not averaged

Results of (type 3) ANOVA between categories indicate that in the word condition there is a main effect of Category: F (2, 227) = 14.7; p = 1.027e-06.

FDR adjusted post-hoc Kolmogorov-Smirnov (ks) tests between categories highlight that, the threshold values at the (A) point are significantly lower than at the points (M) and (Q) : (A) vs. (Q) : ks = 0.35; p = 1.30e-04; (A) vs. (M): ks = 0.37; p = 5.45e-05.

3.3.4.6. Descriptive Statistics of threshold values in A M and Q points in humming condition

Mean values of Threshold Estimates in A M Q categories only in humm condition						
Category	Mean	S.D.	Var.	Min	Max	Obs.
A	16.05059	11.513036	132.55000	1.12	46.20	85
M	15.34065	9.919344	98.39339	1.12	47.18	92
Q	15.31833	10.883614	118.45305	2.10	44.24	84

Table 3.3.15 Descriptive stats of threshold values in humming condition - not averaged

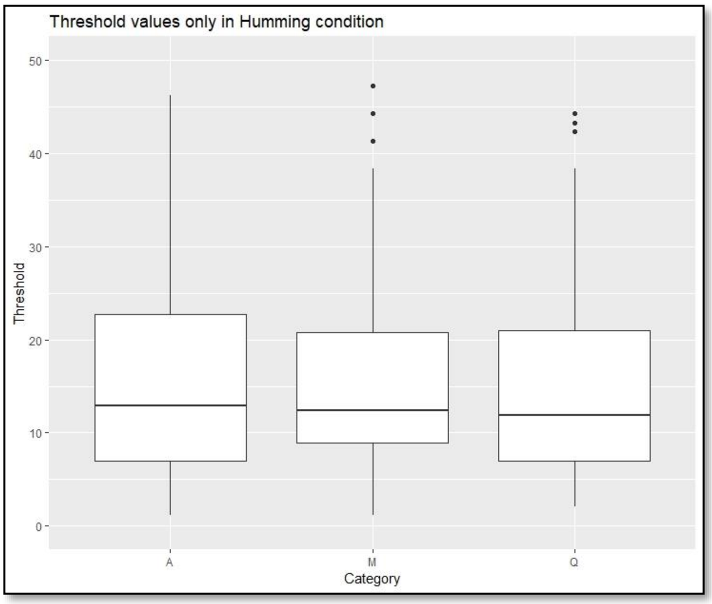


Figure 3.3.11 Threshold values in A M and Q points in humming condition - not averaged

Results of (type 3) ANOVA between the categories do not reveal any effect, signaling that the threshold va lues are equal in all 3 categories (points).

The observation of non-averaged data highlights that there is an effect related to the presence of segments, that lowers of about 5 Hz the threshold level at the point (A) with respect to points (M) and (Q) that is not present in the humming condition. These data suggest that affirmative sentences could be easier to discriminate when segments are available then questions.

3.3.4.7. Raw threshold estimates aggregated per condition, independently from category (A) (M) and (Q)

Mean values of Threshold Estimates in word and humming condition (no category distinction)						
Category	Mean	S.D.	Var.	Min	Max	Obs.
word	18.38504	12.22448	149.4379	2.10	48.16	230
humm	15.56467	10.73266	115.1899	1.12	47.18	261

Table 3.3.16 Descriptive stats of threshold values in word and humming conditions

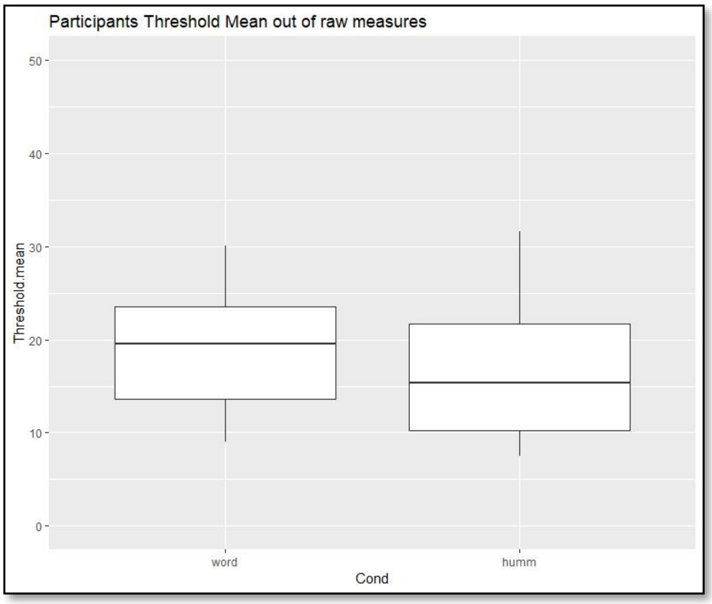


Figure 3.3.12Thresholds in word vs. humming conditions

Paired T-Tests of Mean Threshold Estimates in word and humming conditions (raw measures)				
Mean Threshold Estimates	t	p.value	df	C.I.
	2.6554	0.0161	18	0.5758 4.9395

Table 3.3.17 T-test of threshold values in word and humming conditions

3.3.4.8. Thresholds averaged within participant in point (A) (M) and (Q)

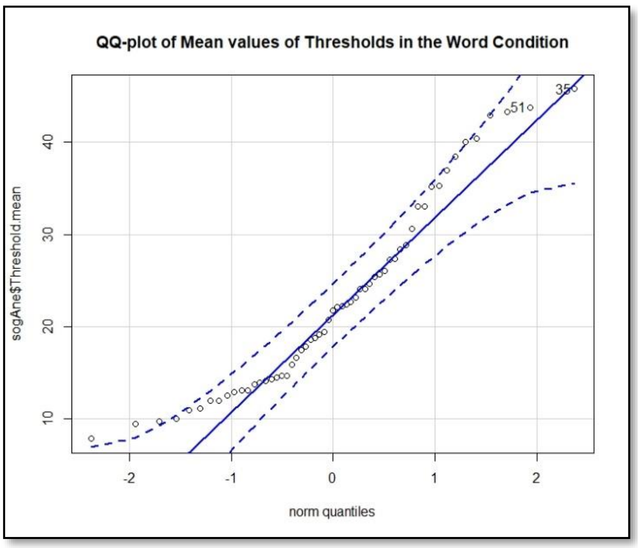
In the following section, I aggregate within participant the measures obtained for each category, obtaining a mean value per subject per category in each condition. Participants for whom only one measure per point was collected are kept. Since the presence of a single observation is not related neither to a specific condition or point, the removal of these data would significantly affect the size of the sample. The measures are kept, aware that the aggregated data comprehending could be noisier.

nsub	Cond	Category	mean	s.d.	Var.	min value	max value
1	word	A	15.62	10.94	119.76	5.04	30.52
1	word	M	19.74	6.74	45.46	9.94	24.64
1	word	Q	30.03	25.64	657.39	11.90	48.16
2	word	A	9.55	8.75	76.64	3.08	24.64
2	word	M	10.72	3.28	10.76	5.04	12.88
2	word	Q	8.96	5.14	26.41	2.10	14.84
4	word	A	6.41	2.65	7.01	3.08	9.94
4	word	M	9.94	5.49	30.09	2.10	14.84
4	word	Q	10.92	4.54	20.65	5.04	16.80
5	word	A	7.98	1.55	2.40	6.02	9.94
5	word	M	15.82	12.47	155.58	7.98	34.44
5	word	Q	16.15	8.22	67.55	10.92	25.62
6	word	A	16.47	7.22	52.18	10.92	24.64
6	word	M	27.58	20.79	432.18	12.88	42.28
6	word	Q	32.15	11.61	134.78	18.76	39.34
7	word	A	10.14	7.76	60.22	5.04	23.66
7	word	M	24.25	12.66	160.19	10.92	44.24
7	word	Q	25.38	8.44	71.31	14.84	35.42
8	word	A	6.76	2.93	8.56	4.06	10.92
8	word	M	18.76	NA	NA	18.76	18.76
8	word	Q	37.38	5.59	31.21	27.58	41.30
9	word	A	22.44	11.88	141.10	12.88	38.36
9	word	M	40.32	NA	NA	40.32	40.32
9	word	Q	21.70	1.39	1.92	20.72	22.68
10	word	A	8.96	7.81	60.99	3.08	22.68
10	word	M	19.15	6.60	43.51	10.92	26.60
10	word	Q	37.05	6.88	47.38	30.52	44.24
11	word	A	22.68	13.60	184.88	7.00	38.36
11	word	M	19.25	10.99	120.69	11.90	35.42
11	word	Q	24.40	16.66	277.48	11.90	48.16
12	word	A	14.51	6.88	47.38	7.98	21.70
12	word	M	14.84	3.53	12.49	11.90	18.76
12	word	Q	35.42	NA	NA	35.42	35.42
13	word	A	11.12	3.21	10.28	7.00	15.82

13	word	M	42.77	3.46	12.01	40.32	45.22
13	word	Q	19.41	22.35	499.73	6.02	45.22
14	word	A	11.51	6.45	41.59	4.06	21.70
14	word	M	13.66	4.35	18.92	7.98	18.76
14	word	Q	11.31	4.93	24.30	5.04	17.78
15	word	A	8.18	5.34	28.52	3.08	14.84
15	word	M	17.78	7.03	49.46	11.90	29.54
15	word	Q	12.88	6.24	38.90	6.02	20.72
16	word	A	4.84	1.28	1.63	3.08	6.02
16	word	M	20.13	10.50	110.25	7.98	31.50
16	word	Q	25.82	8.53	72.70	17.78	39.34
17	word	A	21.11	13.16	173.16	8.96	35.42
17	word	M	11.70	5.16	26.60	5.04	17.78
17	word	Q	10.14	5.34	28.52	3.08	14.84
18	word	A	11.70	2.34	5.47	8.96	14.84
18	word	M	39.93	7.80	60.79	31.50	48.16
18	word	Q	40.81	4.83	23.37	34.44	46.20
19	word	A	21.11	8.13	66.08	11.90	31.50
19	word	M	30.03	6.48	41.94	22.68	36.40
19	word	Q	32.28	10.65	113.52	23.66	48.16
20	word	A	23.07	14.04	197.17	9.94	42.28
20	word	M	33.95	3.46	12.01	31.50	36.40
20	word	Q	7.00	NA	NA	7.00	7.00

3.3.4.9. Thresholds averaged within participant in points: (A) (M) and (Q) – Word condition

In this section I present the mean threshold values aggregated in points: (A) (M) and (Q) in word condition. By averaging the repeated measures at each point the threshold distribution appears to respect normality, and it is possible to perform a repeated measures ANOVA of the factor Category.



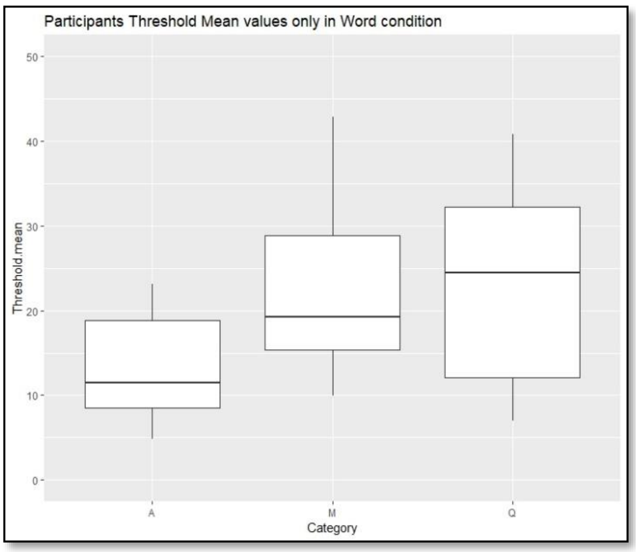


Figure 3.3.13 Normality check and mean thresholds aggregated in (A) (M) and (Q) in word condition

The results of repeated measures ANOVA report a main Category effect F(2, 36) = 8.00; p = 0.001. Post – hoc t-tests, FDR adjusted reveal that the thresholds values within the statement category (A) are significantly lower than the estimated values at the category boundary (M) and within the question category (Q), as reported in Table 3.3.8 below:

Post – Hoc Paired T-Test of individual mean threshold values in each category							
Contrast		t	df	Mean difference	p.value	FDR adjusted p.value	Confidence Interval
1	A – Q	-3.2768	18	-9.7381	0.0042	0.0063	-15.9817 -3.4950
2	A – M	-4.0682	18	-9.2722	0.0007	0.0022	-14.0605 -4.4838
3	M - Q	-0.1591	18	-0.4659	0.8754	0.8754	-6.6195 5.6876

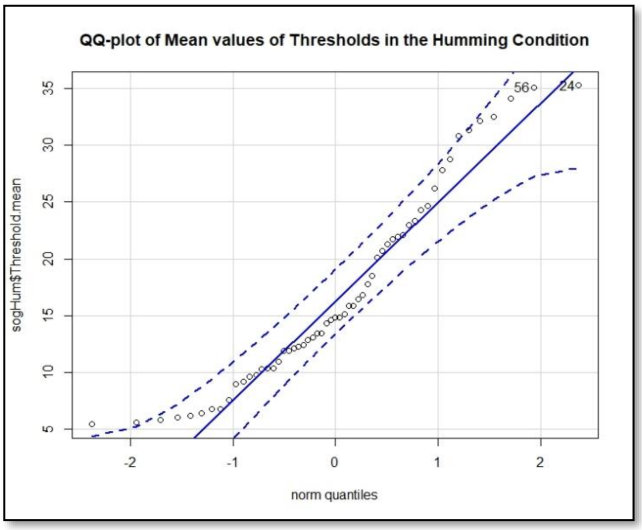
Table 3.3.18 Post-hoc t-tests in word condition

3.3.4.10. <u>Thresholds averaged within participant in points: (A) (M) and (Q) – Humming condition</u>

nsub	Cond	Category	mean	s.d.	var	min	max
1	humm	A	25.23	9.42	88.64	10.92	37.38
1	humm	M	19.74	6.75	45.62	12.88	29.54
1	humm	Q	13.66	4.77	22.76	7.00	18.76
2	humm	A	5.82	1.07	1.15	4.06	7.00
2	humm	M	8.76	1.28	1.63	7.00	9.94
2	humm	Q	8.18	2.72	7.40	5.04	11.90
4	humm	A	13.86	7.88	62.11	6.02	24.64
4	humm	M	11.90	4.75	22.57	7.00	19.74
4	humm	Q	9.35	4.42	19.50	5.04	13.86
5	humm	A	13.86	15.22	231.78	3.08	36.40
5	humm	M	5.43	1.49	2.21	3.08	7.00
5	humm	Q	4.65	2.26	5.09	2.10	7.98
6	humm	A	22.35	13.34	177.99	11.90	37.38
6	humm	M	17.54	8.14	66.19	8.96	28.56
6	humm	Q	31.50	16.63	276.60	19.74	43.26
7	humm	A	11.31	4.88	23.82	6.02	18.76
7	humm	M	15.43	3.96	15.65	11.90	21.70
7	humm	Q	9.94	3.46	12.01	5.04	13.86
8	humm	A	23.27	9.36	87.68	16.80	39.34
8	humm	M	21.11	5.39	29.10	11.90	24.64

8	humm	Q	20.72	12.88	165.83	7.98	38.36
9	humm	A	11.41	7.62	58.10	6.02	16.80
9	humm	M	26.85	9.55	91.16	12.88	34.44
9	humm	Q	34.24	9.18	84.23	21.70	43.26
10	humm	A	7.98	3.46	12.01	4.06	12.88
10	humm	M	15.82	7.23	52.34	8.96	26.60
10	humm	Q	13.37	4.56	20.81	8.96	19.74
11	humm	A	20.33	9.10	82.88	5.04	27.58
11	humm	M	10.92	1.83	3.36	7.98	12.88
11	humm	Q	14.84	10.57	111.79	3.08	30.52
12	humm	A	30.32	12.44	154.82	8.96	40.32
12	humm	M	19.15	3.51	12.29	16.80	24.64
12	humm	Q	10.92	3.92	15.37	7.00	14.84
13	humm	A	14.11	4.48	20.09	8.96	19.74
13	humm	M	9.29	4.93	24.33	4.06	13.86
13	humm	Q	21.95	15.17	230.10	8.96	42.28
14	humm	A	6.61	2.65	7.01	2.10	8.96
14	humm	M	11.12	4.29	18.44	8.96	18.76
14	humm	Q	16.80	5.09	25.93	9.94	21.70
15	humm	A	4.45	3.37	11.33	1.12	8.96
15	humm	M	12.10	2.54	6.43	8.96	14.84
15	humm	Q	5.82	2.97	8.84	2.10	9.94
16	humm	A	5.04	1.55	2.40	3.08	7.00
16	humm	M	9.35	4.68	21.90	2.10	13.86
16	humm	Q	12.49	5.44	29.58	6.02	17.78
17	humm	A	12.49	6.41	41.11	6.02	19.74
17	humm	M	4.84	1.07	1.15	4.06	6.02
17	humm	Q	5.24	0.82	0.67	4.06	6.02
18	humm	A	20.92	11.56	133.69	7.98	39.34
18	humm	M	31.11	15.89	252.39	4.06	44.24
18	humm	Q	29.79	9.68	93.72	23.66	44.24
19	humm	A	27.78	10.36	107.28	15.82	38.36
19	humm	M	8.63	4.95	24.46	1.12	13.86
19	humm	Q	23.66	9.62	92.58	11.90	35.42
20	humm	A	33.13	12.75	162.63	20.72	46.20
20	humm	M	34.05	9.69	93.93	24.64	47.18
20	humm	Q	14.84	NA	NA	14.84	14.84

The mean threshold values aggregated in points: (A) (M) and (Q) in humming condition show overall to respect normality. The repeated measures ANOVA of the factor Category reveal that within the humming condition the factor Category does not cause a significant effect. The discrimination sensitivity within the humming condition shows to be independent from the presence of categories.



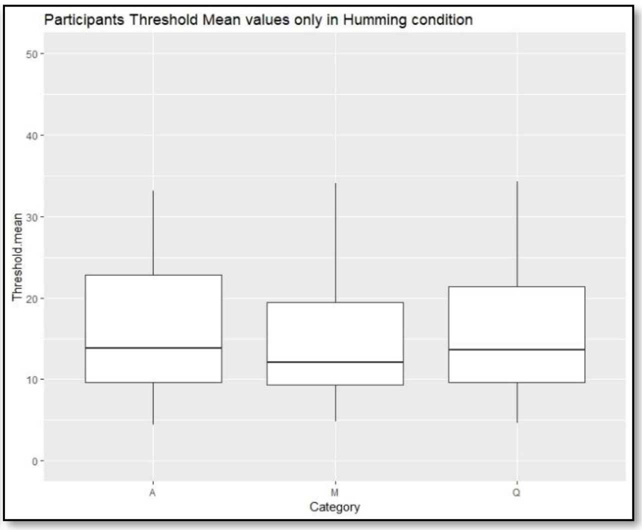


Figure 3.3.14 Normality check and mean thresholds aggregated in (A) (M) (Q) in humming condition

3.3.4.11. <u>Mean values comparison per condition and per category</u>

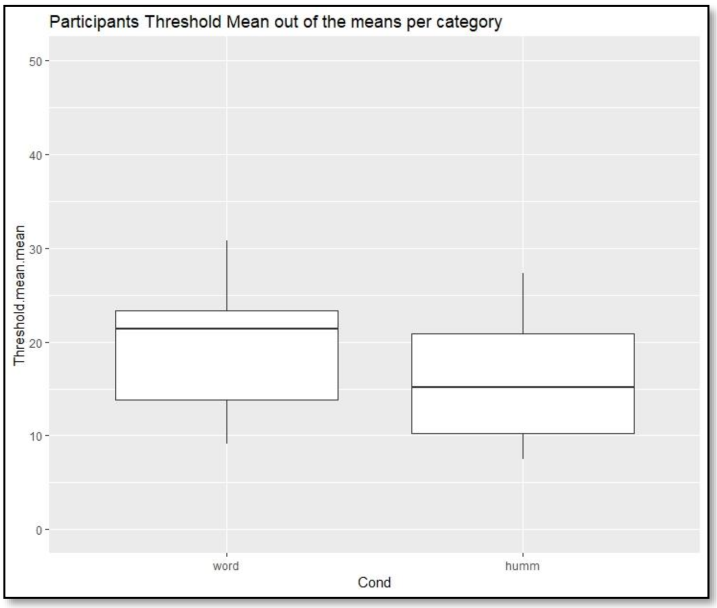


Figure 3.3.15 Mean thresholds aggregated per condition

Mean of participants' mean values of Threshold in word and humming condition (no category distinction)						
Category	Mean	S.D.	Var.	Min	Max	Obs.
word	19.71364	10.274418	105.56366	4.844	42.770	57
humming	15.88333	8.553766	73.16691	4.452	34.244	57

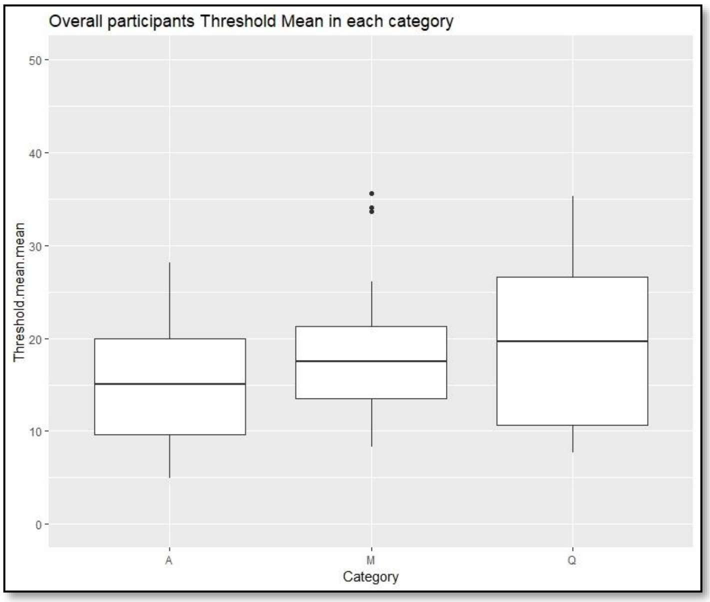


Figure 3.3.16 Mean thresholds aggregated per category

Results of repeated measures ANOVA report a main Category effect F(2, 36) = 3.79; p = 0.03.

Post – Hoc Paired T-Test of individual mean threshold values in each category aggregating the 2 conditions							
Contrast		t	df	Mean difference	p.value	FDR adjusted p.value	Confidence Interval
1	A – Q	-2.2581	18	-4.6503	0.0366	0.0549	-8.9769 -0.3237
2	A – M	-2.3515	18	-4.1852	0.0303	0.0549	-7.9244 -0.4460
3	M - Q	-0.2699	18	-0.4651	0.7903	0.7903	-4.0856 3.1555

Table 3.3.19 Post-hoc t-tests between categories between mean values independently from cond.

Overall the analysis shows that within the humming condition the discrimination sensitivity is independent from the presence of linguistic categories and it is constant across the range of presented stimuli.
Compared to the word condition, without considering the presence of categories, the discrimination sensitivity within the humming condition is lower than with the presence of segments.
When observing the impact that categories have on perception, it observable that within word condition the stimuli within the statement category are better discriminated than the stimuli within the question

category and positioned on the category boundary. When merging the data and observing the role of categories independently from the condition then the discrimination sensitivity appears to be lower within the statement category and possibly shows the tendency to linearly model in function of frequency values of the final boundary tones. Lower frequencies appear to be better discriminated than the higher frequencies. The data overall show to support the presence of continuous perception in absence of segments, and an impact of pragmatic categories on perception when segments are able that is more in line with the proposal of a perceptual magnet within the statement category compared to the proposal of the classic CP paradigm.

3.3.4.12. Categorical Perception of Intonation

In this section I investigate the relation within each participant between the results obtained in the categorization task, and the results obtained in the discrimination task.
In particular I am interested in observing whether the relationship between the logistic curve and the threshold estimates highlights the presence of Categorical Perception of intonation.
To do so, I compute Linear Models (LMs) of the slope and midpoint values in each condition, in function of the participants' mean threshold estimate in each category (A, M, Q). This analysis allows to verify whether the presence of a steeper categorization curve affects the discrimination rate in the second task, and whether the higher impact on discrimination capabilities is found as an overall effect in the 3 categories or specifically in the M category, signaling that CP boosts the discrimination sensitivity in the midpoint perceptual area.

3.3.4.12.1. Linear Modeling of slope values and mean threshold values at points (A) (M) and (Q) in word condition

When modeling slope values within the word condition with the mean threshold estimates in the same condition, I find no significant correlations in the category (A): slope.w ~ Mean.Thresh.Aff.w (beta0 = 0.1190, SE = 0.0154; beta = -0.0016, SE = 0.0011; t (16) = .1.473; p = 0.16; r = -0.3457) and (Q): slope.w ~ Mean.Thresh.Que.w (beta0 = 0.0862, SE = 0.0150; beta = 0.0005, SE = 0.0006; t (16) = 0.895; p = 0.384; r = - 0.2184). When modeling the correlation between slope value and the threshold estimate in the category (M) I find a tendency, reported in the Table 3.3.20 and Figure 3.3.17 below.

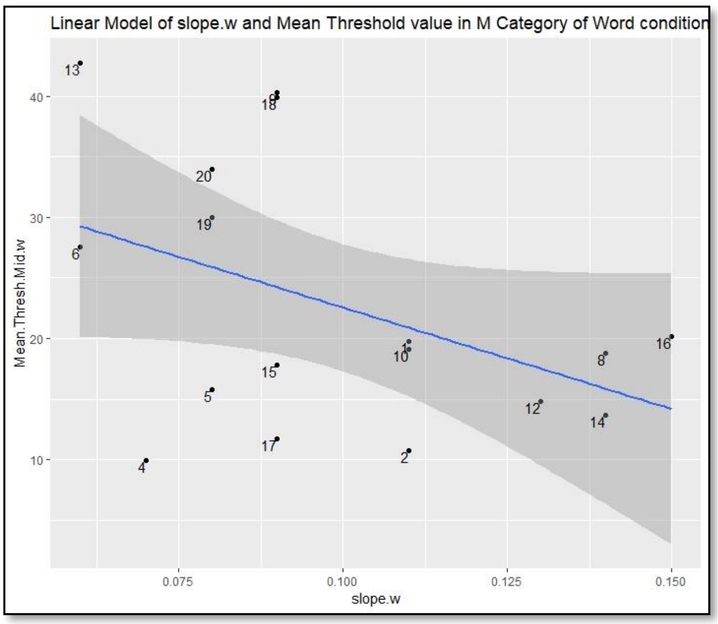


Figure 3.3.17 LM of slope values and (M) mean threshold in word condition

Linear Model of slope values in word condition and threshold in M category								
Linear Model	df	beta0	SE.	beta	SE.	t (beta)	p.value (beta)	Corr. value
slope.w ~ Mean.Thresh.Mid.w	16	0.1238	0.0146	-0.0011	0.0006	-1.9150	0.0735	-0.4318

Table 3.3.20 LM of slope values and (M) mean threshold in word condition

3.3.4.12.2. Linear Modeling of slope values and mean threshold values at points (A) (M) and (Q) in humming condition

The results of LMs of slope values in the humming condition and thresholds estimates in the same conditi on highlight, also in this condition, no significant correlations in the category (A) (beta0 = 0.1048, SE = 0.01 35; beta = 0.0004, SE = 0.0007; t (16) = 0.583; p = 0.568; r = 0.1442) and (Q): (beta0 = 0.1223, SE = 0.0131; beta = -0.0007, SE = 0.0007; t (16) = -0.953; p = 0.335; r = -0.2318).
LM of slope values and the threshold estimate in the category (M) in the humming condition highlight a te ndency, reported in the table 3.3.21 and Figure 3.3.18 below.

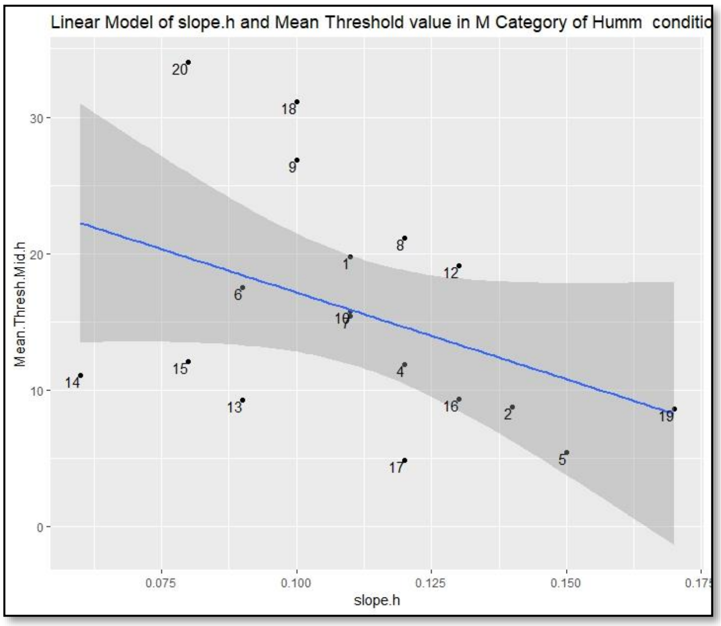


Figure 3.3.18 Figure 3.3.17 LM of slope values and (M) mean threshold in humming condition

Linear Model of slope values in Humming condition and threshold in M category								
Linear Model	df	beta0	SE.	beta	SE.	t (beta)	p.value (beta)	Corr. value
slope.h ~ Mean.Thresh.Mid.h	16	0.1321	0.0130	-0.0013	0.0007	-1.783	0.0936	-0.4071

Table 3.3.21 LM of slope values and (M) mean threshold in humming condition

3.3.4.12.3. Linear Modeling of midpoint values and mean threshold values at points (A) (M) and (Q) in word condition

LMs of the relationship between the midpoint and the threshold estimates in the word condition display n o significant correlation in the (A) category (beta0 = 54.9177, SE = 7.0917; beta = 0.5371, SE = 0.5042; t (1 6) = 1.065; p = 0.303; r = 0.2573) and in the (M) category (beta0 = 59.65784, SE = 7.21692; beta = 0.09487 , SE = 0.28817; t (16) = 0.329; p = 0.746; r = 0.08202881). LM of midpoint values and the threshold estimat e in the category (Q) in the word condition highlight a tendency, reported in the table 3.3.22 and Figure 3. 3.19 below.

Linear Model of midpoint values in Word condition and threshold in Q category								
Linear Model	df	beta0	SE.	beta	SE.	t (beta)	p.value (beta)	Corr. value
midpoint.w ~ Mean.Thresh.Que.w	16	72.1028	6.2916	-0.2464	0.2464	-1.81	0.0891	-0.4122

Table 3.3.22 LM of midpoint values and (Q) mean threshold in word condition

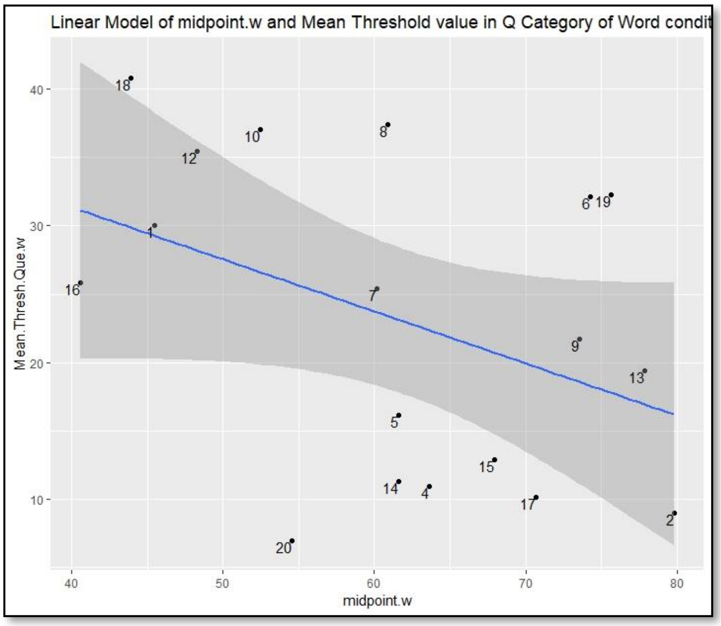


Figure 3.3.19 LM of midpoint values and (Q) mean threshold in word condition

3.3.4.12.4. Linear Modeling of midpoint values and mean threshold values at points (A) (M) and (Q) in humming condition

LMs of the relationship between the midpoint and the threshold estimates in the humming condition displ ay no significant correlation in the (A) category (beta0 = 73.2173, SE = 5.8898; beta = -0.2918, SE = 0.3203; t (16) = -9.11; p = 0.376; r = -0.2221) and in the (Q) category (beta0 = 66.0182, SE = 5.9449; beta = 0.1567 , SE = 0.3256; t (16) = 0.481; p = 0.637; r = 0.1195). LM of midpoint values and the threshold estimate in th e category (M) in the humming condition highlight a tendency, reported in the table and plot below.

Linear Model of midpoint values in Humming condition and threshold in M category								
Linear Model	df	beta0	SE.	beta	SE.	t (beta)	p.value (beta)	Corr. value
midpoint.h ~ Mean.Thresh.Mid.h	16	77.5348	5.750 7	-0.5751	0.3250	-1.77	0.0958	-0.4046

Table 3.3.23 LM of midpoint values and (M) mean threshold in humming condition

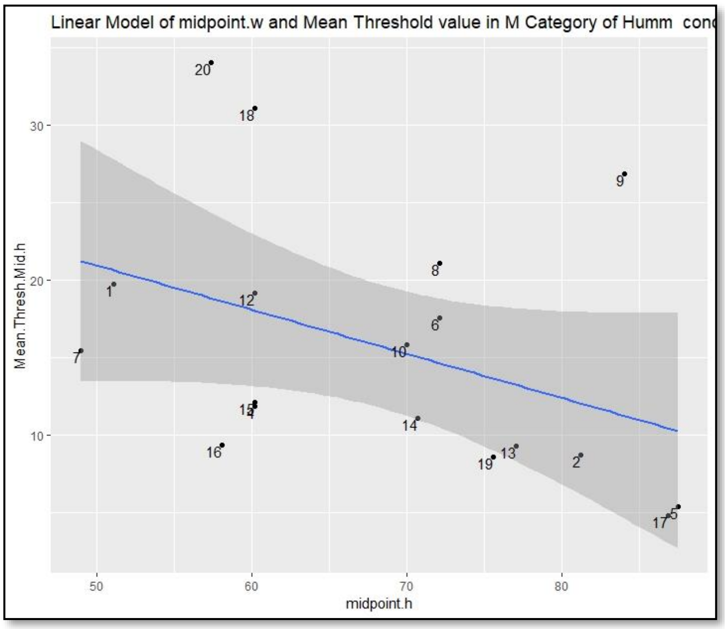


Figure 3.3.20 LM of midpoint values and (M) mean threshold in humming condition

3.3.4.13. Categorical Perception of Intonation – Single subjects' threshold- trends:

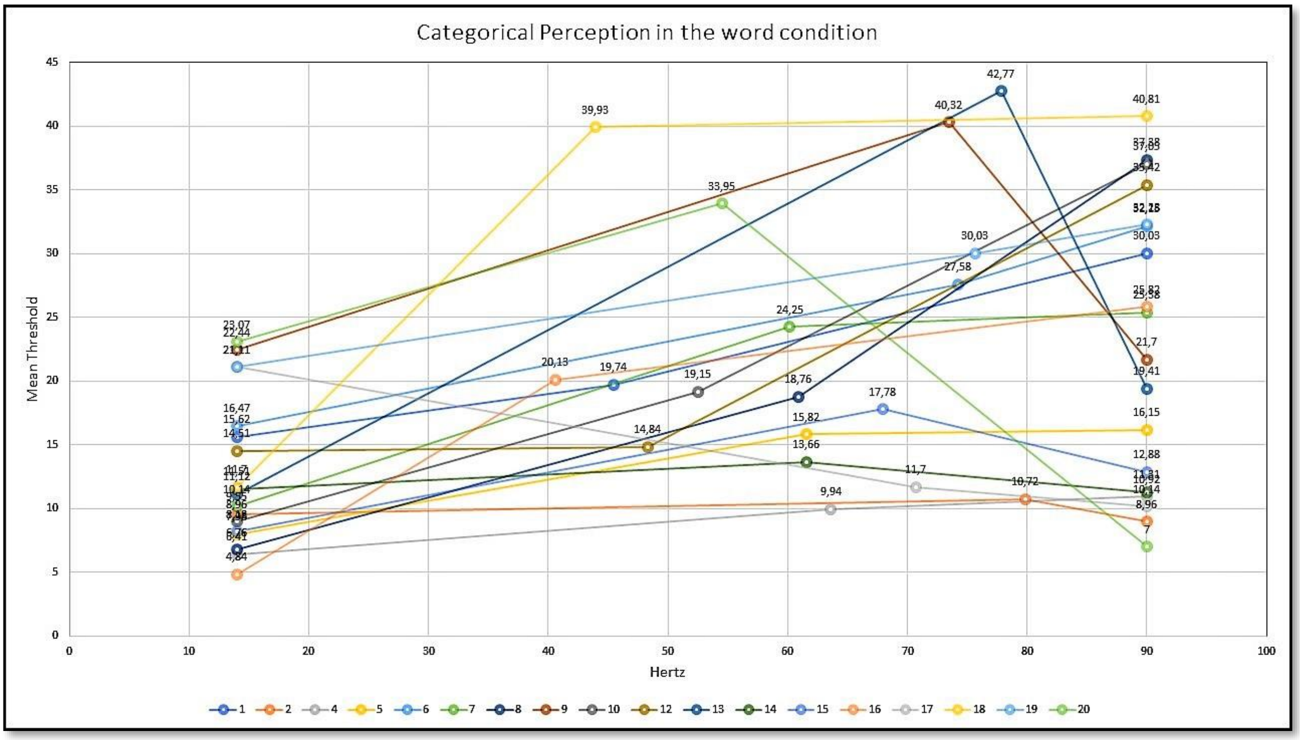


Figure 3.3.21 Single subjects thresholds trends in word condition

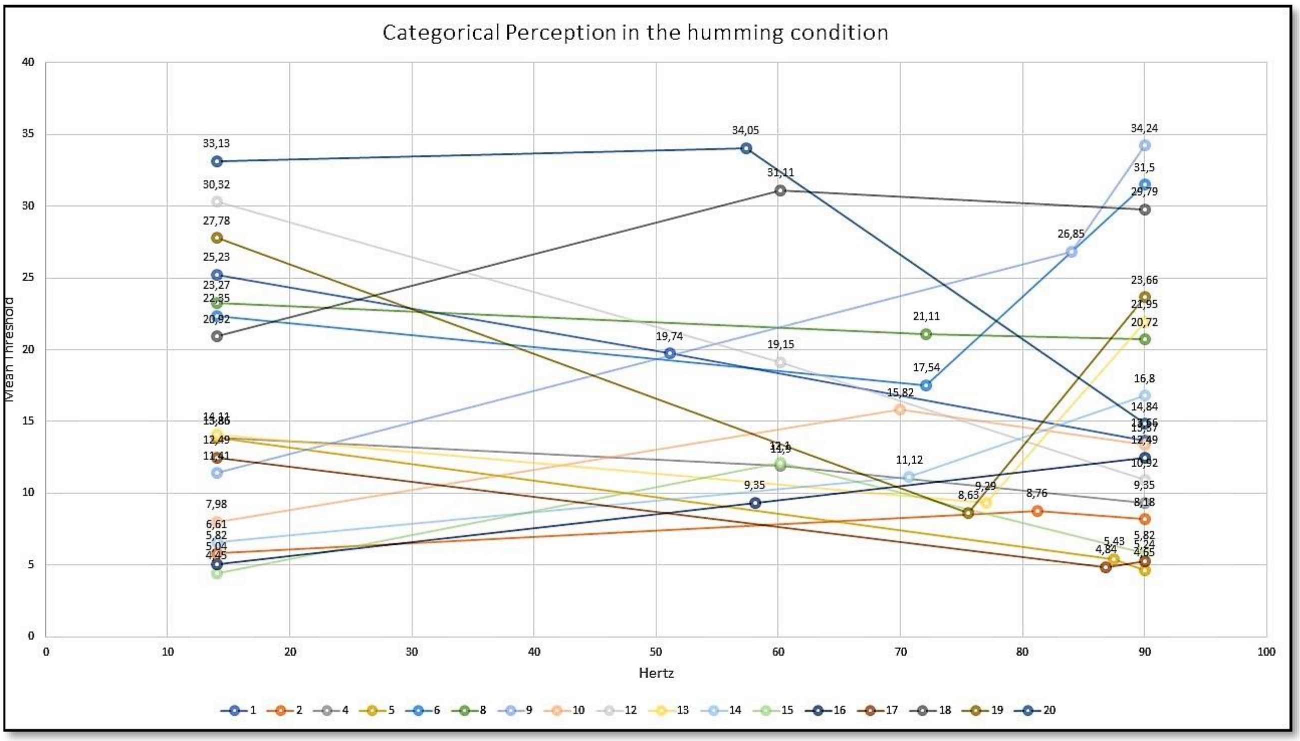


Figure 3.3.22 Single subjects thresholds trends in humming condition

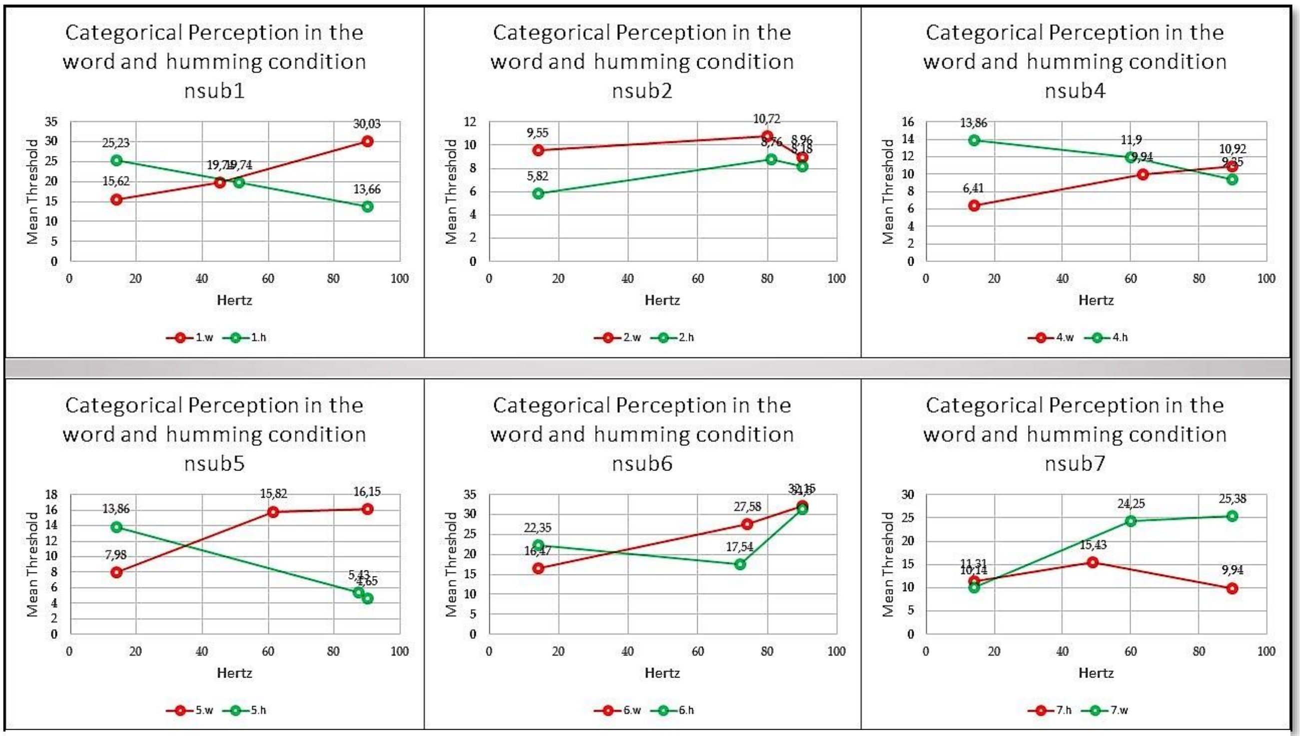


Figure 3.3.23 Single subjects' comparison of thresholds trends in word and humming condition (1)

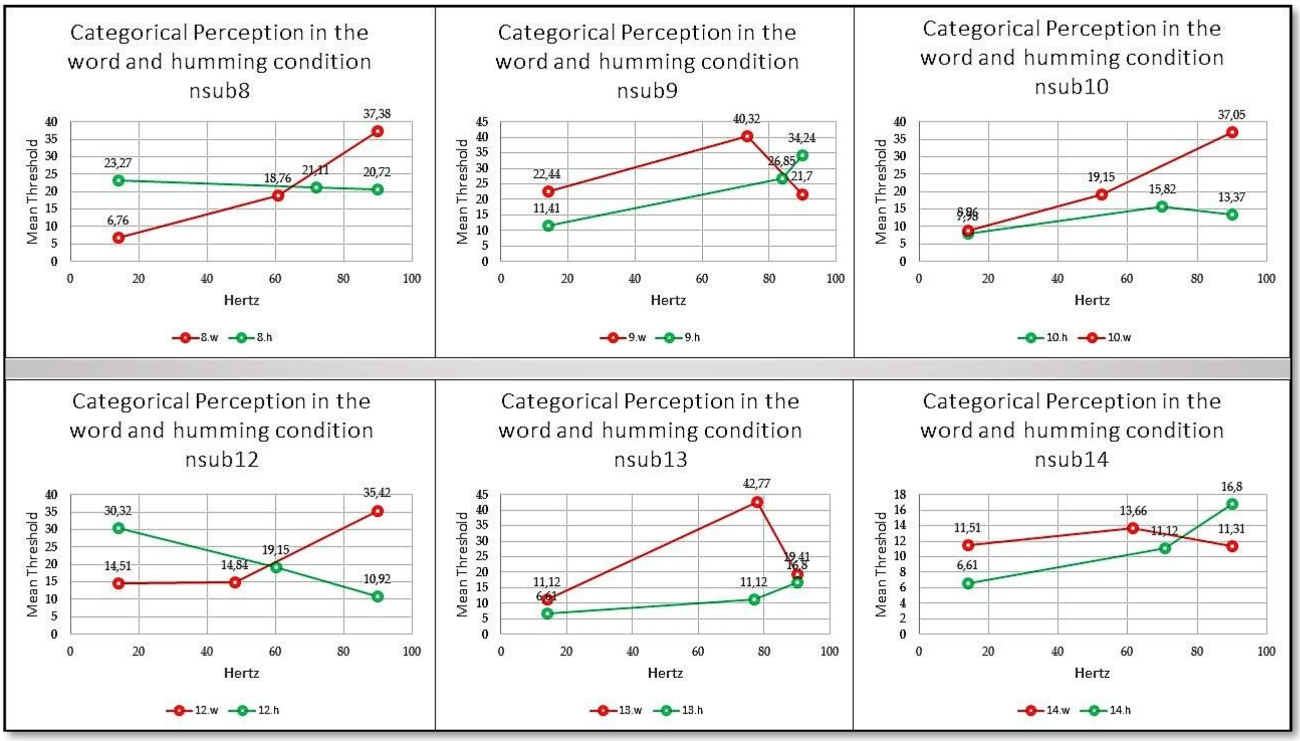


Figure 3.3.24 Single subjects' comparison of thresholds trends in word and humming condition (2)

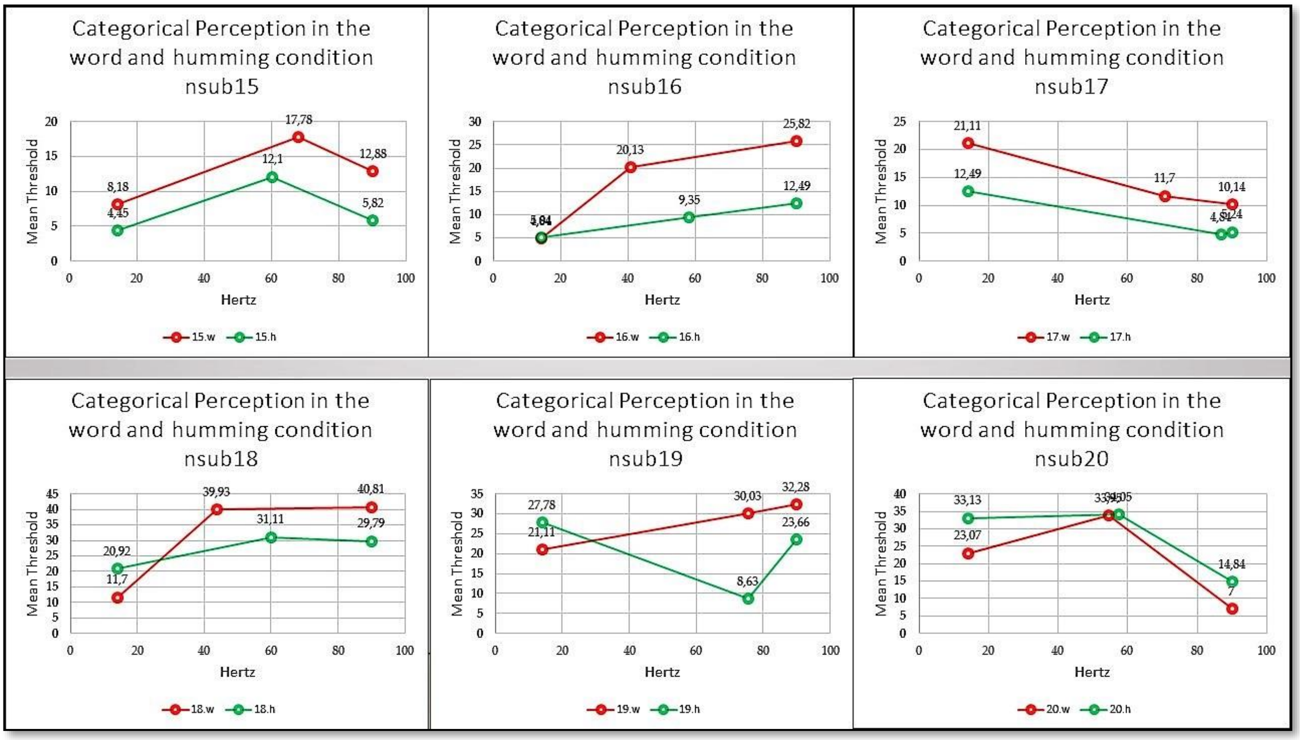


Figure 3.3.25 Single subjects' comparison of thresholds trends in word and humming condition (3)

3.3.5. Correlated Measures of Individual Variability and Discrimination's thresholds

The Principal Component Analysis (PCA) of AQ and IRI variable performed in paragraph 3.3.3 yielded the following the results.

	Dim1	Dim2	Dim3	Dim4	Dim5	Dim6	Dim7	Dim8	Dim9
S.D.	1.7194	1.2622	1.1353	1.0109	0.8681	0.7782	0.6629	0.4192	0.4068
Prop of Variance	0.3285	0.1770	0.1432	0.1135	0.0837	0.0673	0.0488	0.0195	0.0183
Cumulative Prop	0.3285	0.5055	0.6487	0.7622	0.8460	0.9133	0.9621	0.9816	1.0000

Table 3.3.24 PCA dimensions

In the following analysis I select Dim2 and Dim3 (Dim1 despite explaining the largest proportion of variance, does not discriminate well the sample of participants on the basis of the provided variables) explaining a total proportion of variance of about the 32%.

The following biplot in Figure 3.3.24 illustrates the distribution of variance within the sample, and how well each variable is represented combining the two independent components (the longer the arrow, the better the representation).

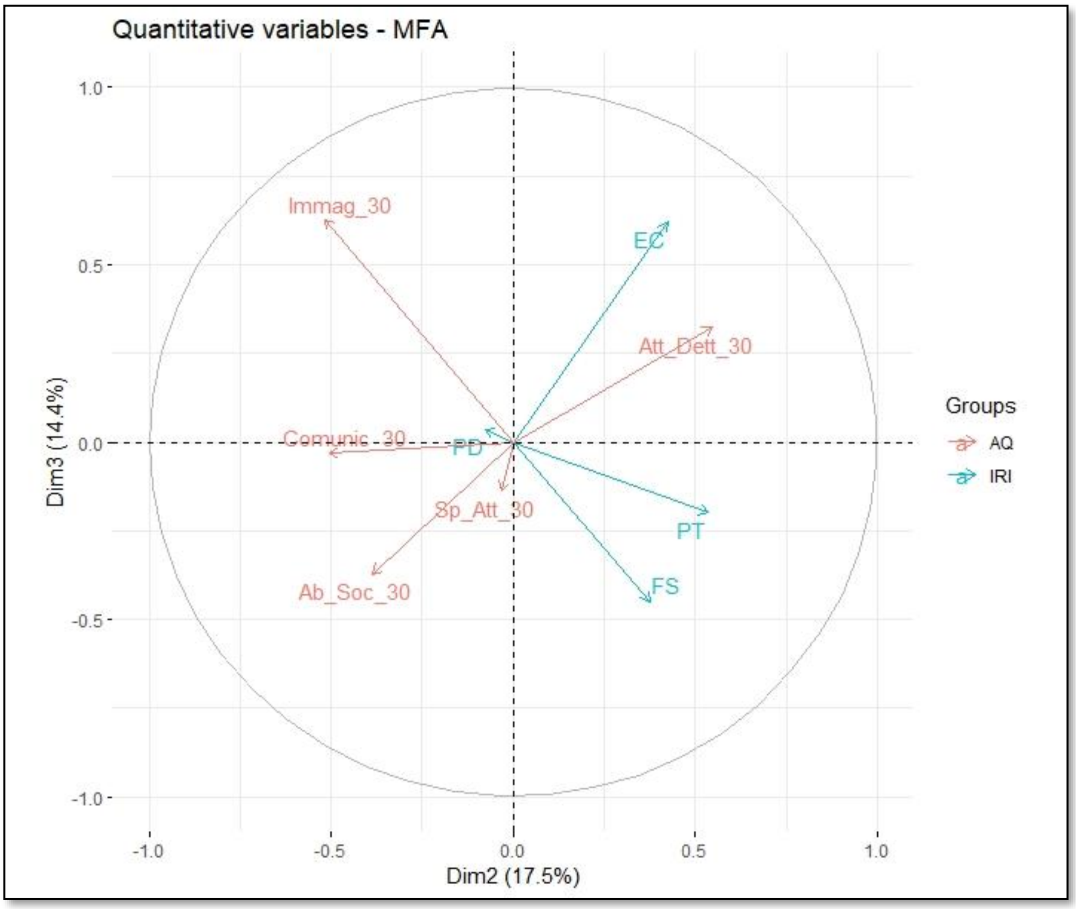


Figure 3.3.24 Biplot of Dim2 and Dim3 of PCA of AQ and IRI variables

In the following Figure 3.3.25 the composition of each of the two selected dimensions is represented

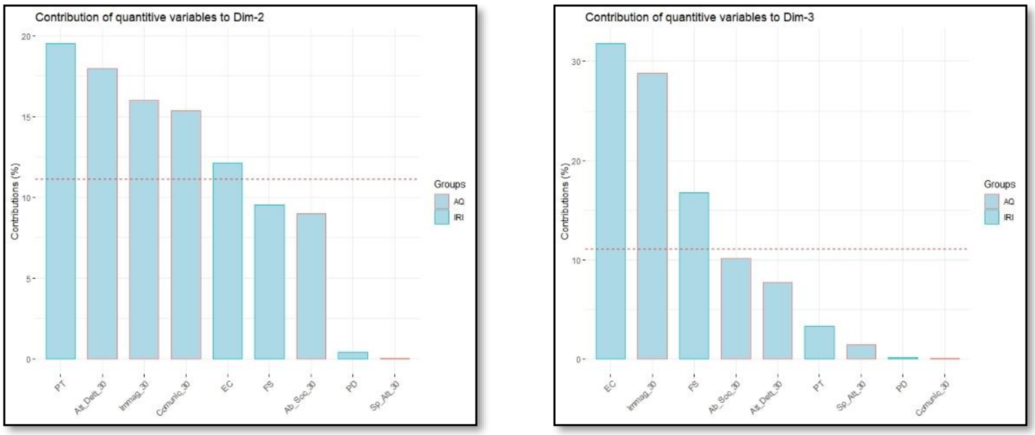


Figure 3.3.25 Composition of Dim2 and Dim3

Since the PC analysis is computed adopting 9 variables, in the case each variable contributed in equal manner to the Dimension composition, it would explain a proportion of variance of 100/9 = 11.1%. Variables contributing more than the threshold value of 11.1% have a higher weight within the Dimension and therefore connotate it. Dim2 therefore orients the participants in function their score in the variables PT, Att_Dett_30, Immag_30, Comunic_30, and EC. Dim3 orients the participants in function of their scores in the scales EC, Immag_30 and FS. The most suitable dimensions explain the highest proportion of variance, discriminate each variable within the biplot representation and cumulative cover the highest number of relevant variables. In this way participants can be further sub-divided in clusters that differ between each other in the composition of the represented variables in the dimensions. (see Chapter 4, for a richer explanation of the procedure of cluster composition).

From Figure 3.3.26 to 3.3.30, the position of each of the 20 participants along Dim2 and Dim3 in function of the variables that mostly differentiates subjects within the sample are plotted. Higher values in the Perspective Taking (PT) and Empathy Concern (EC) scale (ceiling value = 28) indicate a stronger presence of the Theory of Mind construct. They are assumed to facilitate the pitch processing. Higher values in the AQ variables, indicating an atypical Attention to Details (Att_Dett_30), the Resistance to Attentional Switching (Sp_Att_30) indicate the tendency to show a modulation of attention similar to the one of individual diagnosed with ASC (ceiling value = 30). Higher values also in the social and communicative scales of the AQ test, also are interpreted as a potential impairment in the managing of social situations (Ab_Soc_30 and Comunic_30; ceiling values = 30). Higher values in the scale Imag_30 signal a difficulty in the imagination of fictional scenes and negative correlate with higher values in the FS (IRI test, ceiling value = 28) signaling a predisposition to imagine fictional scenes. Finally, the Personal Distress scale (ceiling value = 28) signals the tendency of the individual to suffer from a great amount of stress and anxiety in critical situation, and higher values of PD are assumed here to be a disadvantage in the performance of the discrimination task.

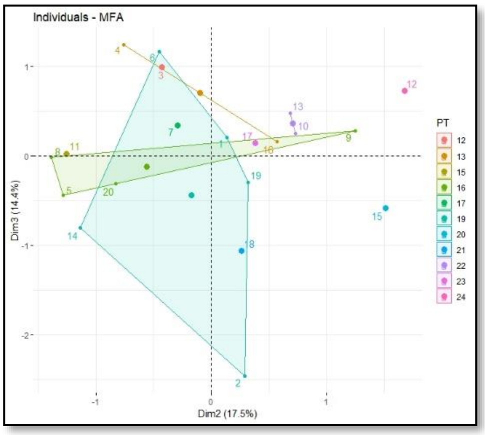


Figure 3.3.26 PT

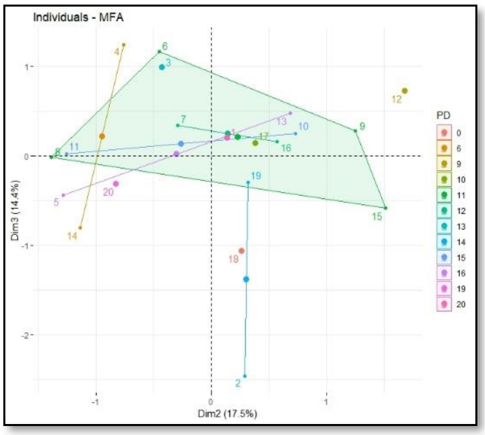


Figure 3.3.27 PD

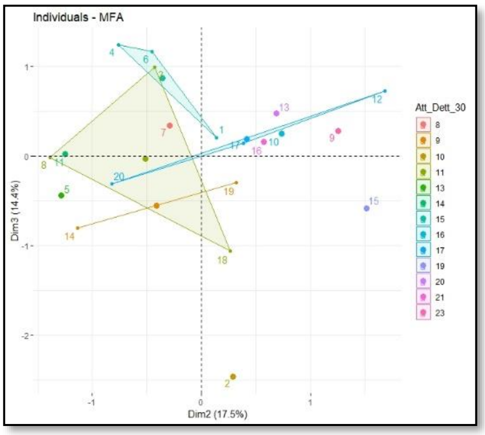


Figure 3.3.28 Att_Dett_30

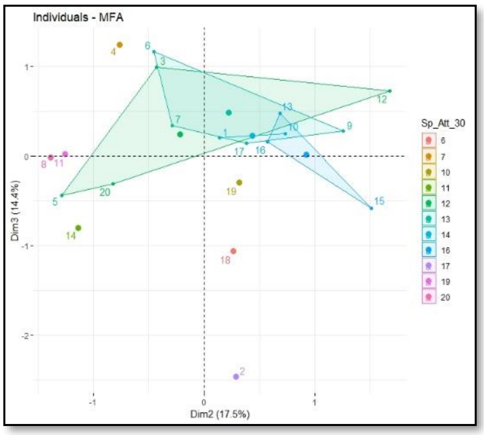


Figure 3.3.29 Sp_Att_30

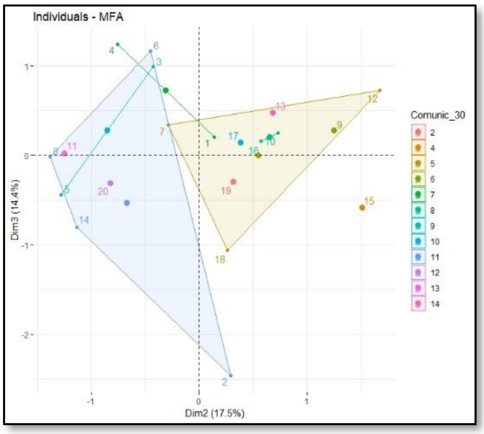


Figure 3.3.30 Communic_30

The results of the PCA highlight as expected that the sample is sub-divisible adopting a main criterion that divides participants having higher values in the variables indicating a social impairment (Ab_Soc_30, Communic_30) and difficulties in imagination (Immag_30) on one side, and having higher values in the variables assumed to represent an advantage for the interpretation of the pragmatic function of discourse (EC, PT, FS).

Interestingly the variable Att_Dett_30 correlates in the direction assumed to indicate an expected advantage of that specific group of participants.

3.3.6. Clusters formation

Results of the PCA analysis are taken as input for Hierarchical Clustering on Principal Components (HCPC). The aim is to observe whether different clusters of the sample, on the basis of AQ and IRI variables can account for perceptual discrimination differences of final rising Boundary Tones. In particular, the interest is to investigate whether higher values of variable assumed to advantage the pragmatic interpretation of pitch (EC and PT) actually correlated with lower discrimination thresholds, in Word other Humming conditions, and a stronger presence of Categorical Perception, indicated by lower threshold values in the perceptual area surrounding the Midpoint.

The results of the HCPC provide the following clusters' subdivision that will be aggregated in two main clusters with the aim to maximize the differences given the limited number of participants. Individuals positioned in Cl1 and Cl5 display the strongest contraposition in the values of the social variables: namely strongly signaling the presence of a hypothetical impairment (higher values in Immag_30 and Comunic_30 and lowest in EC and PT); on the contrary participants in Cl3 and Cl4, display the exact opposite pattern. In order to achieve a similar number of observations in each group, the participants in the center will be excluded, since they are not well represented by the analysis.

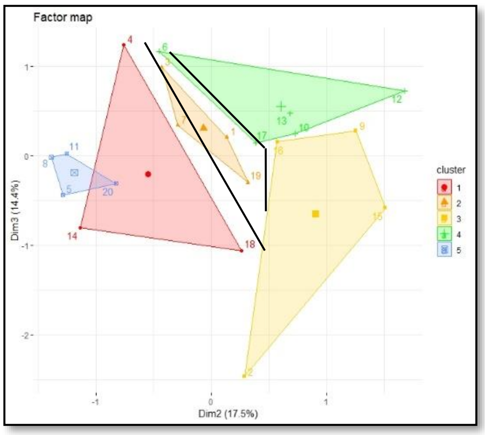


Figure 3.3.31 Cluster formation on the basis of AQ and IRI variables

The resulting final composition is, thus, reported in the following Tables 3.3.25 and 3.3.26:

Cluster 1

nsub	FS	EC	PT	PD	Ab_Soc_30	Sp_Att_30	Att_Dett_30	Comunic_30	Immag_30	cl
4	7	17	13	6	5	7	15	7	11	1
5	18	15	16	16	16	12	13	9	12	5
8	19	18	16	11	13	20	11	11	14	5
11	18	18	15	15	12	19	14	13	12	5
14	12	13	19	6	11	11	9	11	9	1
18	10	12	21	0	8	6	11	5	5	1
20	18	21	16	19	21	12	17	12	9	5

Table 3.3.25 Cluster 1 - Discrimination Threshold Exp

Cluster 2

nsub	FS	EC	PT	PD	Ab_Soc_30	Sp_Att_30	Att_Dett_30	Comunic_30	Immag_30	cl
2	28	16	19	14	13	17	10	11	3	3
6	13	24	19	11	12	13	15	11	13	4
9	22	19	16	11	5	13	23	6	8	3
10	20	25	22	15	14	14	16	8	10	4
12	14	25	24	9	7	12	17	5	8	4
13	21	25	22	16	11	16	20	14	10	4
15	24	20	20	11	10	16	19	4	7	3
16	23	21	13	12	9	16	21	8	8	3
17	20	20	23	10	11	13	17	10	12	4

Table 3.3.26 Cluster 2 - Discrimination Threshold Exp

Threshold's average values of participants in Cluster 1

nsub	Cond	Category	Threshold.mean	Threshold.sd	cluster
4	word	A	6.412	2.647814193	1
4	word	M	9.94	5.48566617	1
4	word	Q	10.92	4.544073063	1
5	word	A	7.98	1.549516053	1
5	word	M	15.82	12.47336362	1
5	word	Q	16.14666667	8.218767142	1
8	word	A	6.755	2.926357235	1
8	word	M	18.76	NA	1
8	word	Q	37.38	5.586859583	1
11	word	A	22.68	13.59694819	1
11	word	M	19.25	10.98591219	1
11	word	Q	24.395	16.65759787	1
14	word	A	11.508	6.44866808	1
14	word	M	13.664	4.34969884	1
14	word	Q	11.312	4.929312325	1
18	word	A	11.704	2.339717932	1
18	word	M	39.928	7.797007118	1

18	word	Q	40.81	4.834225205	1
20	word	A	23.072	14.04172781	1
20	word	M	33.95	3.464823228	1
20	word	Q	7	NA	1
4	humm	A	13.86	7.880727547	1
4	humm	M	11.9	4.75072626	1
4	humm	Q	9.352	4.415441088	1
5	humm	A	13.86	15.22420879	1
5	humm	M	5.432	1.486243587	1
5	humm	Q	4.648	2.256129429	1
8	humm	A	23.268	9.364001282	1
8	humm	M	21.112	5.394452706	1
8	humm	Q	20.72	12.87746352	1
11	humm	A	20.328	9.103983743	1
11	humm	M	10.92	1.83341212	1
11	humm	Q	14.84	10.5731055	1
14	humm	A	6.608	2.647814193	1
14	humm	M	11.116	4.294144851	1
14	humm	Q	16.8	5.092229374	1
18	humm	A	20.916	11.56233886	1
18	humm	M	31.108	15.88688516	1
18	humm	Q	29.785	9.680859122	1
20	humm	A	33.13333333	12.75255791	1
20	humm	M	34.048	9.691600487	1
20	humm	Q	14.84	NA	1

Table 3.3.27 Threshold values in Cluster1

Threshold's average values of participants in Cluster 2

nsub	Cond	Category	Threshold.mean	Threshold.sd	cluster
2	word	A	9.548	8.754422882	2
2	word	M	10.724	3.279707304	2
2	word	Q	8.96	5.139163356	2
6	word	A	16.47333333	7.2236925	2
6	word	M	27.58	20.78893937	2
6	word	Q	32.15333333	11.60931235	2
9	word	A	22.435	11.87850019	2
9	word	M	40.32	NA	2
9	word	Q	21.7	1.385929291	2
10	word	A	8.96	7.809314951	2
10	word	M	19.152	6.595916919	2
10	word	Q	37.05333333	6.883293785	2
12	word	A	14.51333333	6.883293785	2
12	word	M	14.84	3.53344025	2
12	word	Q	35.42	NA	2
13	word	A	11.116	3.205663738	2

13	word	M	42.77	3.464823228	2
13	word	Q	19.41333333	22.35459982	2
15	word	A	8.176	5.340775225	2
15	word	M	17.78	7.032823046	2
15	word	Q	12.88	6.23668181	2
16	word	A	4.844	1.277763671	2
16	word	M	20.132	10.50018667	2
16	word	Q	25.816	8.5265632	2
17	word	A	21.112	13.15903188	2
17	word	M	11.704	5.157817368	2
17	word	Q	10.136	5.340775225	2
2	humm	A	5.824	1.073536213	2
2	humm	M	8.764	1.277763671	2
2	humm	Q	8.176	2.719389637	2
6	humm	A	22.35333333	13.3414442	2
6	humm	M	17.535	8.135574145	2
6	humm	Q	31.5	16.63115149	2
9	humm	A	11.41	7.622611101	2
9	humm	M	26.845	9.547668127	2
9	humm	Q	34.244	9.177531258	2
10	humm	A	7.98	3.464823228	2
10	humm	M	15.82	7.2347633	2
10	humm	Q	13.37	4.561651748	2
12	humm	A	30.324	12.44252707	2
12	humm	M	19.152	3.506154589	2
12	humm	Q	10.92	3.92	2
13	humm	A	14.105	4.48200476	2
13	humm	M	9.286666667	4.932558498	2
13	humm	Q	21.945	15.16891009	2
15	humm	A	4.452	3.366410551	2
15	humm	M	12.096	2.536667105	2
15	humm	Q	5.824	2.972487174	2
16	humm	A	5.04	1.549516053	2
16	humm	M	9.352	4.679435863	2
16	humm	Q	12.488	5.438779275	2
17	humm	A	12.488	6.411327476	2
17	humm	M	4.844	1.073536213	2
17	humm	Q	5.236	0.819926826	2

Table 3.3.28 Threshold values in Cluster2

3.3.7. Comparison of Threshold average values in Cluster1 and Cluster2

On the basis of the computed HCPC, individuals in Cluster1 are expected to present a higher impairment in the decoding of the pragmatic function of communication, compared to individuals in Cluster2.
I present here the results of an explorative analysis seeking to observe whether an impairment in the adoption of the perspective of the interlocutor, as well as in the understanding of the general principles guiding communication, as well as difficulties in imaging fictional scenes correlates with an advantage in discriminating pitch movements with and without segments.

The results of the t.tests investigating a main difference between the two Clusters in the discrimination of minimal differences in pitch height in word condition vs. in humming condition, report a non-significant effect. Average threshold values in word condition of participants in Cluster and Cluster2 are substantially equal. A slight non-significant difference in the expected direction is present within the humming condition, but far from reaching significance ($p < 0.4$).

When observing the presence or absence of CP within condition and within cluster, only the word condition within cluster2 displays a tendency ($p < 0.1$) for the factor CATEGORY, indicating threshold values significantly different in (A) (M) and (Q) categories $F(2,16) = 3.27$; $p = 0.06$.

Overall the results do not indicate that the measured covariate variables related to the processing of pragmatic aspects of communication significantly impact on discrimination capabilities.

3.4. Discussion

The results of the Categorization task, with respect to the results of Experiment 1, show that the linguistic activation granted by the word condition presented in 1st place has a rather lower effect on the perception of the humming category – within a linguistic modality – compared to the effect it appeared in sequence 1 of Experiment 1.
Slope values across conditions show the tendency to correlate, but not reaching significance. In addition, slope values in the humming condition could not be predicted on the basis of the obtained values in the word condition. The mixed models' comparison highlighted a stronger presence of individual variability, beginning with the word condition.
On the other hand, midpoint values, that in Experiment 1 appeared to be more bound to the physical properties of the stimuli and affected by the order of presentation, showed in this Experiment 2 to behave similarly, averaging reaching the half – range of frequency and displaying the presence of a fixed effect within the mixed models analysis.
On one side, this highlights that the sequence in Experiment 1 had much stronger training potential with respect to the presentation of only 2 conditions with very opposite features pertaining to segmental information.
This is relevant to keep in mind when performing categorical perception tasks with more than 2 conditions. If it is not the aim of the study, the presence of an ordered sequence strongly supports the formation and recognition of categories even adopting informationally poorer signals.

On the other side, the partial replication of the results is encouraging. They showed to provide relatively stable information on the possible functional interpretation of slope and midpoint values, once a logistic function is selected, and the resulting curve is fitted on the participants' responses in CP tasks.
Commonly the categorization task is used only to signal the presence of the two categories, but commonly it is not adopted to reach a more in-depth modeling explanation. These two experiments showed that their behavior could be correlated in linguistic conditions, but mainly refers to two conceptually different perception strategies.
The slope value seemed to represent better a top-down activation of an existing category that in the presence of repeated stimuli (such it is in a longer sequence containing more than one linguistic condition as in experiment 1) can be adopted to categorize stimuli not fully belonging to speech. On the contrary, the midpoint value seemed to be much more related to a bottom-up categorization process that strongly relies on the previously heard stimuli and transfers the generalization strategy from the adjacent condition.
This functional interpretation of slope and midpoint values represents a possible source of improvement for future research in the field.

The results of the discrimination task overall highlighted a substantial absence of the Categorical Perception as classically intended (Liberman, 1961). Out of 18 participants only 2, displayed significantly lower discrimination thresholds values associated with their category boundaries, and interestingly the effect pertained the humming category and not the linguistic one (nsub6 and nsub19).
On the contrary, rather unexpectedly at the group level, the discrimination rate clearly showed a trend linearly increasing from the statement category to the question category, displaying a facilitatory effect of the category statement in the linguistic condition.
This, in light of the presented literature in the introduction sections of Chapter 1 and Chapter 2 could be motivated supporting 3 main claims:

In line with the evidence reported by House (1990), the presence of segments, increasing the complexity of the signal, shows to affect perception probably at the level related to the presence of a Critical Bandwidth (CB) (Von Békésy & Wever, 1960). In the data, this is observable in the analysis investigating groups' comparison only in function of the word and humming condition. The absence of segments lowers the discrimination threshold of about 3 Hz when compared to the word condition, and in both cases the presence of formants transitions in both the linguistic and non-linguistic conditions does not allow a discrimination rate averagely lower than 15 Hz. The psychoacoustic literature on perception of pure tones commonly reports a minimum threshold sensitivity for pure tones in the orders of units, below 10 Hz (Kingdom & Prins, 2010). This supports the theory that the presence of spectral information reduces the overall sensitivity to F0 movements, slowly changing in time, and that the lowest impact is found at lower frequencies compared to high.

The second implication of these results is that, despite a different experimental paradigm and manipulation, they support the proposal of Van Heuven & Kirsner (2004) and the studies of Schneider (2012) that highlight communicative and perceptual properties of statements opposed to questions. In particular, both studies support the view that: (i) a more categorical shift in perception occurs in the shift a from a clearly descending contour, associable to the non-marked declination effect pattern, to a rising contour in the opposite direction; and that (ii) a perceptual magnet, provided with a prototypical perceptual representation, is present in statements and not in questions. From these results, it appears that this is confirmed and that it has a relation with the communicative function of speech since the effect is not present in the humming condition alone.

The third implication of these results is that the category boundary did not display any perceptual and discriminatory advantage when compared to the discrimination rate of the stimuli within-category. On the contrary, participants not showing a linear trend seem to show more consistently higher threshold estimates in the category boundary area than the contrary.
This could be related to the adopted stimuli. Overall the final rise of 140 Hz representing the question category could represent a too much higher frequency range, forcing the participants to adopt a different perception strategy rather than the normally adopted one during speech comprehension. The choice of 140 Hz was motivated as a choice that led with no doubt to the recognition of the question category. Future research could replicate this experiment by adopting a lower range of frequency modulation. A related technical, aspect that could have affected or biased the results is the possibility that the adaptive procedure calculated the estimates with a possibly too high error rate, given the number of observations that had to be discarded at the beginning of the analysis. Also, the fact that I kept single observations for some participants, not having averaged values, might have increased the variance in the question category resulting in a higher mean value.
There could also be another artifact related to the experimental material used in the tasks, scaled in Hz and not in semitones, that could have facilitated the discrimination at lower frequencies compared to higher ones. Given the adoption for the first time of an adaptive procedure within a CP paradigm on intonation, further experiments could help improve the quality of the estimates and of the stimuli.

In third instance, it could be possible that the adaptive procedure, estimating a minimum threshold, in truth investigates a different aspect of the discrimination sensitivity, compared to the standard ABX procedure presenting stimuli at a fixed range. The discrimination function peak, in truth, does not indicate that in the presence of a category boundary participants could perceive a smaller pitch difference, compared to within-category boundaries. On the other hand, the task seems plausible in light of the presence of a Perceptual Magne Effect. If within category, it is still possible to judge a stimulus as a better category-instance with respect to other instances of the same category and the perceptual magnet attracts the most similar members to it, then, it is reasonable to think that far from a perceptual a magnet it should be possible to discriminate a more fine-grained signal. This question remains open for further research.

A further open question is the tendency for thresholds values estimated at the category boundary (M) to decrease as slope values in the categorization task increases. This phenomenon, contrary to what has been previously discussed, could support the presence of a perceptual mechanism related to the categorical perception. It signals that the more a participant displays a sharp transition between categories the lower will be the minimum threshold at that point. The tendency appeared only at the point (M) and not within category, therefore it could be possible that if modulating the overall range of frequency not to be so strongly affected by the Critical Bandwidth effects, it could be possible for participants to more clearly recognize the question category, and as dependent effect to improve in the discrimination rate at the Category Boundary.

In final instance, the analysis of the discrimination capability in function of the collected covariate measures related to the presence of an assumed advantage or disadvantage in disambiguating the pragmatic interpretation of discourse, revealed that as expected participants polarized in two subgroups displaying on one side higher values in the variables indicating a disadvantage (Cluster1) and on the other one, higher values in the variables indicating a possible advantage (Cluster2). Despite the polarization of the covariate measures, the resulting clusters did not significantly differ in their average threshold values, in word vs. humming conditions, or within condition across categories.

This possibly indicates that a pragmatic profile of the participant might impact only on the identification/generalization process, but it does not directly affect lower sensory mechanisms. Given the fact that the two experiments combine acoustic, linguistic, and individual cognitive advantages or disadvantages in the decoding of the pragmatic function of pitch in a new fashion compared to previous experiments, and the modelling of individual variability is still performed with a limited number of observation, it will be necessary to collect more data in order to provide a more certain explanation. Nonetheless, since the pragmatic profile impacted more on the categorization task in Experiment 1, and as it will be seen, it shows to be relevant in the following Chapter, this might suggest that the complex nature of intonation, composed of both low and higher cognitive processes is modulated by different individual variabilities at different levels, depending on the task.
A very fine-grained discrimination ability might not be required to compute linguistic processing of pitch; therefore, it is plausible to think that the pragmatic profile of a participant impacts more on the higher linguistic computations that occur later on in the processing, significantly after a pure sound discrimination phase.

In conclusion, it is important to highlight, that the collected sample of participants is highly homogeneous, belonging to the same macro-regional area, representing only the female population, without musical and foreign languages' training. Such a sample has been collected on purpose in order to exclude variability that could come from those factors and to collect baseline data for future research, not affected by other variables. Nonetheless, it could also be possible that individual variability emerges when starting considering all those other levels. For example, differences between female and male participants, and significantly different personal experiences of each individual.

3.5. Conclusions

Taken together the two psychoacoustical experiments presented in Chapter 2 and Chapter 3 highlighted that the perceptual nature of intonational categories, especially in function of pitch height, is far more complicated to model, compared to the analysis of segments.
In line with what previously presented by Schneider (2012), intonational categories indeed display properties of half-closed categories and multidimensional nature. The presence of segments and the recognizability of word- and sentence- meaning significantly contribute to the linguistic, pragmatic function that pitch can have in speech. Pitch is not invariantly perceived independently from the presence or absence of segments. On one side, the results of these two psychoacoustic experiments support the current neuroscientific data indicating the presence of different neural networks devoted to the perception of pitch with linguistic function, and without linguistic function, e.g. emotional prosody vs. linguistic prosody (Wildgruber, Ackermann, Kreifelts, & Ethofer, 2006; Wildgruber et al., 2004). The pragmatic function that pitch can assume in speech represents, therefore, a refined integration of several cues spanning over different areas of spoken language, ranging from phonetic features to the recognition of meaning.
On the other side, these results possibly support (i) the proposal of Hickok & Poeppel (2004, 2007)of a mechanism relying on the presence of bilateral multi-parallel routes, carrying also information on the time-

scale resolution of signals emerging from different types of processing, and (ii) their proposal of a partially shared initial stage perceptual mechanism processing speech and non-speech that then diverges at later stages in function of linguistic experience. This could at the end explain the presence of the overall converging evidence in CP paradigms' experiments involving the perception of Boundary Tones (see paragraph 2.2.), showing a general ability to identify distant opposite intonational categories, but then reporting non-converging evidence on the discrimination rates within categories and at the boundaries.

Overall, considering the presented literature, I believe the results can support the proposal of a paradigmatic shift towards the focus on the categorization process itself (Holt & Lotto, 2010; Heald et al., 2017) from paradigms simply adopting and seeking Categorical Perception. The data overall supported that participants in the condition of firstly perceiving linguistic stimuli seem to be generally capable of finding an association rule when switching to not fully linguistic stimuli, that grants them the ability to recognize sentence modality also in the humming condition. Nonetheless, the ability to recognize the presence of two opposite pragmatic functions appeared not to be related with enhanced discrimination sensitivity associated to the presence of a boundary tone and the ability to recognize categories related to sentence modality showed to be perceptually acquired in the course of the experimental sequence starting with stimuli in which segments were poorly recognizable. This seems to indicate that the identification of a category is not strictly linked to the presence of speech. The category can also be acquired in the absence of segments, what changed was the strategy adopted to form it (emerging in slope and midpoint values). Therefore, segments and thus language seems not to be the cause of CP, and the constraints of CP instead segments impact on the selected cognitive mechanism adopted to solve a CP task.

In second instance, the results obtained in these two experiments, framed within the classical CP paradigms results are in line with previous experiments, since the category recognition is largely confirmed both in Experiment 1 and 2, but the discrimination results in Experiment 2 do not support the presence of a strict CP and indicate a possible perceptual advantage of lower frequencies discrimination in linguistic modality – pitch within the statement category is discriminated to a finer level compared to pitch modulations over category boundaries and within the question category.
Nonetheless, the two experiments add evidence pertaining the methodological level, since they both are the first indicating that the psycholinguistic logistic functions within the identification task can be modeled to investigate the perceptual response at the variation of stimulus type (conditions from 1 to 4) as well as possible correlated measures of a different nature (results of questionnaires). None of the presented studies in the introductory sessions ever attempted to provide more details and explanation on the interpretation of the parameters of the logistic identification function, nor, more relevantly ever attempted to model both parameters of the identification function, with parameters of the discrimination function, and external factors.
These two psychoacoustical experiments represent therefore a first attempt to improve the research within the intonation field and to start including individual variables in the perceptual models, and a more unified perspective at theoretical level.

Thirdly, the evident different nature of the results involving higher order linguistic representations, emerging in the identification/categorization task, compared to the evidence related to lower perceptual, sensory mechanisms, emerging in the discrimination task pushes for a discussion of the current issues in the theory of intonation.

The time-scale of the mechanisms necessary to process the linguistic function of intonation, the higher effort to retain the incoming signal in memory, the required processing of co-occurring segmental information, as well of the need for multiple repetitions and for stable spectral information, and the absence of a clear-cut categorical perception, indicate that the intonational categories are not directly available to the perceiver in the time-scale of segments, nor can be presumably directly mapped on the basis of motor representations of the movement of the vocal tract. The whole reported evidence, as well as the presented literature on pitch perception, strongly indicates that the recognition of the categories is achieved by integrating several higher order levels of abstract information. The sensory processing alone is not sufficient.
This signals that the current direction that the Second Language Acquisition models are taking, firmly relying on models directly descending from approaches related to the Motor Theory of Speech Perception, has to be evaluated reformulating or re-discussing the steps through which categories are perceived, recognized and acquired in L2. The presented models in paragraphs from 1.5.2. to 1.5.4. probably capture the perceptual nature of segments, affecting pitch perception, but might not capture the comprehensive abstract nature of the prosodic category itself. The communicative power of the intonational instrument does not lie in an extremely fine-grained discriminatory skill at the lower sensory level, but rather in the efficiency of the generalization and identification process. The categorization strategy can be trained and improve only considering information coming from signals of a diverse nature, and in the most efficient cases switches in function of the type of stimulus. For these reasons, the models of L2 acquisition should begin to consider the integration process, and should begin to model how the phonetic information can assume an abstract representation that integrates with meaning and the further pragmatic interpretation, such as modality, considering the advantages provided by a perspective such as the Polysp, and maintaining the lightest possible modeling structure.

In the following Chapter, I will address the mapping of intonation with specific aspects of the discourse structure in native and second language speakers. The Chapter will focus on the role of referential accessibility and information packaging within the discourse, adopting a psycholinguistic perspective and a neurophysiological technique. In addition, it addresses how prosody maps with the discourse structure, and how the presence of different mapping rules present in the native language of the perceivers affects the pitch and discourse processing in L2.

Electroencephalography (EEG) will allow considering both the response to the physical properties of the stimuli in the auditory domain, as well as higher-order cognitive operations considering the discourse level. The EEG experiment links to the presented psychoacoustical experiments as it if were a fine-details continuation over time of the auditory responses investigated in the first chapters. The use of more articulated stimuli, involving a pair of meaningful sentences, as well as the EEG technique increases the complexity of the debate.

Nonetheless, in all 3 experiments, the crucial aspect remains the understanding of the mapping between sound and higher-order representations, assuming the semantic and pragmatic status. Finally, it is crucial in all 3 experiments to understand the perceptual strategy of the participants and the way they make use of the perceived pitch modulation and co-occurring segmental information. In all 3 experiments, the underlying hypotheses are that the variability at the individual level in responding either to the auditory domain or to a more efficient use of contextual information increases the quality of the perceptual processing.

Chapter 4
The processing of German Pitch Accents by Italian learners of German

4.1. Outline

In this Chapter, I focus on the processing of the integration between intonation and discourse information structure in native (L1) and second language (L2) speakers of German, native speakers of Italian and in an immersion condition at the time of the experiment. I present a perceptual study conducted with the use of a neurophysiological measure, Electroencephalography (EEG), that allows observing the online processing of stimulus with a refined time-resolution.

The experiment is constructed as a replication and adaptation to L2 speakers of German, of the paradigm originally adopted by Schumacher & Baumann (2010). In the original paradigm, the authors investigated, in German, the interplay between the cognitive activation of a referent (see, Chapter 1: the referential approach) and intonation. The original idea behind the choice to adopt the same paradigm, was to use the results of Schumacher & Baumann (2010) as baseline for comparison of L1 and L2 processing, focusing on the processing of the prosodic marking of semi-active referents, and specifically addressing the processing of Deaccentuation, given the ongoing debate on plasticity and the assumed typological difference present between Romance and West Germanic languages (see Chapter 1, paragraph 1.2.7.).

Nonetheless, in the course of adaption of the experimental material to be easily perceived by non-native speakers, the new adapted stimuli have been recorded varying the phonetic implementation of an experimental condition. In particular, the enhancement of the clarity of the speech and a slower speech tempo compared to naturally uttered speech in German of the adopted speech style of the new stimuli, resulted in a systematic alteration of one of the 3 conditions, that was originally intended to correspond to the H* accent. The newly produced stimuli in this condition presented a much higher and steeper peak than the original ones, leading to the possible interpretation of the experimental condition, not as simply, completely new and inactive information, as it was meant, but adding a possible Narrow Contrastive Focus (CF) interpretation. The new stimuli have been found to be more appropriately labeled as L+H*, an accent associable to a CF in both of the languages involved in the study (cit).

By choosing to keep the newly recorded experimental condition, the experiment considers both the prosodic marking of the information status of a referent, in function of its cognitive degree of activation (Referential approach), and the prosodic manipulation of a focus type within a focus-background framework (Relational approach).

I present in the following sections a brief recap of the topic of interests, fully addressed in Chapter 1, and of the open questions pertaining to the processing of intonation in L2. Then I specifically report the available literature on the processing of discourse structure in the visual and in the auditory domain at the neurophysiological level, focusing on the Event-Related Components (ERP) that have been found to be related with the prosodic marking of discourse structure.
I, then, introduce the specific predictions pertaining to the processing Deaccentuation and Focus marking in German L2, by native speakers of Italian, given the debate on plasticity, addressed in Chapter 1.

In final instance, I introduce the correlated measures collected to investigate the presence of external and internal factors in L2 processing of pitch modulation that are adopted to perform the specific analysis of the presence of possible clusters within the sample of L2 learners. This separated investigation has to aim to delineate some of the relevant covarying factors impacting on L2 acquisition, as well as the role of specific individual characteristics of the participants determining a more native-like or less native-like L2 processing of prosodic marking.

4.2. Introduction

As introduced in Chapter 1, a referent can assume at least 3 states within discourse: new, accessible, and given. A referent is considered *new* when it is mentioned for the first time within a discourse model, cognitively shifting from a status of complete inactivation to full activation. It is considered *given* when it refers to a fully identifiable extra-linguistic entity, already mentioned in discourse (contextually salient), that does not add new information to the existing mental representation of the discourse model, and that is anchorable to a specific antecedent in discourse. The level in between, corresponding to the *accessible* information, is associated with a referent sharing a partial representation with already mentioned referents within discourse that is inferable from the context. It represents a cognitive semi-active state in which the linguistic term is newly introduced, but the informational content carried by the term is anchorable through inference to previously activated information(Baumann, 2005, 2006; Chafe, 1994).

In German, referents in discourse have been consistently found to be prosodically marked in function of their degree of accessability (see Chapter 1 paragraph 1.2.5) (Baumann & Grice, 2006; Baumann & Hadelich, 2003; Grice et al., 2006; Kohler, 1991; Roehr & Baumann, 2010; Röhr & Baumann, 2011). In particular, it has been reported a preferential marking with H* of completely new referents and the consistent lack of prosodic prominence – Deaccentuation – in cases of textually given referents. Accessible referents, on the other hand, are reported to be preferentially marked through the falling contour H+L*(Baumann & Grice, 2006).

In the past 10 years, neurophysiological research has begun to investigate the cognitive processing of referential accessability of referents in discourse, adopting online measures such as Electroencephalography (EEG) firstly in the visual domain, and recently, also addressing the function of prosody in the construction of the discourse model (e.g. Baumann & Schumacher, 2012; Schumacher & Baumann, 2010).

Nonetheless, to author's knowledge, the processing of the integration of prosody and information status of a referent in second language (L2) learners has not yet been addressed.

In Chapter 1, I have presented the open issues pertaining the nature of an intonational category and the challenges that the field of Second Language Acquisition (SLA) has to face when developing hypotheses on the acquisition process of intonation, since the growing body of literature strongly relies on models originally developed addressing the perception and acquisition of segments.
The most accepted models of SLA of segments, the Speech Learning Model (Flege, 1995) and the Perceptual Assimilation Model (C. Best, 1995; C. T. Best & Tyler, 2007) that are the base of the developing theory of L2 acquisition of prosody, the L2 Intonation Learning Theory (Mennen, 2015) highlight a strong role of the phonetic features during acquisition. Briefly summarizing a general overview of the literature (see Chapter

1 paragraph 1.5.), the mentioned models predict that the acquisition in L2 is guided by a perception modality that is filtered from the perceptual categories present in the native language, that represents a template to which incoming new features are assimilated or not assimilated (Flege, 1995; Best & Tyler, 2007; Gili Fivela, 2012; Mennen 2015). A successful acquisition is predicted when the L2 features are recognized as linguistic features, opposed to the possibility of non-linguistic processing (Best & Tyler, 2007), sufficiently diverging from the available phonetic representation of the L1 categories.
Nonetheless, the perceptual dimension of segments is tightly connected with the motor representation of the movements of the vocal tract, and this is assumption has been historically deeply rooted in the formulation of perception theories as in the Motor Theory of Speech Perception (Liberman & Mattingly, 1985) and the realist approach (Best, 1995; Galantucci et al., 2006), that are respectively at the base of the Categorical Perception paradigm (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967) and of the Perceptual Assimilation Model (Best, 1995).

Intonation, on the other hand, is modeled across languages assuming a central role of the phonological representations, that are assumed to represent linguistic discrete, recognizable categories (Grice & Baumann, 2007) and also assuming a determining role of phonetic implementation(Mennen, 2004, 2007) but that is less clear whether it entails continuous, quasi-categorical, or categorical perception (See Chapter 1 and 2 for a full discussion of the nature of prosodic categories and their perception modalities),
Moreover, intonation is primarily generated by the movement of the vocal chords, and its perception is guided by the decoding of the Fundamental Frequency (F0), and by its interaction with the spectral information generated by the vocal tract articulating segments, as we have seen in Chapter 2 and 3. At a higher level, it is integrated with the meaning of the referent to which it aligns, it is processed in function of its metrical and syntactical position within an utterance, and it has to be integrated with the information decoded in previously perceived utterances.
The function of phonetic features and of abstract phonological representations assumes an even more relevant role investigating the acquisition of intonation. This is particularly relevant, on one side, because phonetic features strongly interact with the presence of segments generating Categorical or Quasi Categorical perception, and in second instance, because they have been found to be the elements that mostly convey universal meanings attached to intonation (Gussenhoven, 2002).
On the contrary, the abstract discrete intonational phonological entities, tightly bound to the structure of the utterance, represent the linguo-specific categories and it is not yet clear how universal mechanisms interact with linguo-specific constraints and structures and how this is reflected in the L2 acquisition.
Therefore, it is evident that the literature on L2 acquisition presented in Chapter 1 paragraph 1.5. has not yet addressed the full spectrum of relevant variables and cannot yet clearly predict the L2 processing of intonational categories.
The processing of intonation in L2 can be affected by the available phonetic and phonological categories of the native language, as predicted by the available models. Nonetheless, very few it is known on how the perception of intonational categories, affected by the presence of native language's representations, is further integrated at a higher level with the word and sentence meaning, in the construction of discourse.

For these reasons in this Chapter, I present a study, adopting the EEG technique, in which I adapt to L2 learners an experimental paradigm developed by Schumacher & Baumann (2010).

Schumacher & Baumann (2010) specifically constructed an experiment manipulating a particular type of bridging inference: The whole-part relationship (see Chapter 1.2.3). As described in Chapter 1 paragraph

1.2.3., the “part” element within a whole-part relationship represents accessible semi-active information, previously and partially activated by the “whole” concept, implicitly activating a set of related referents. Semi-active referents in whole-part (and scenario) relationships introduce in the discourse entities that can be properly anchored to an antecedent and update the mental representation with a new referent. As previously described, discourse-updating (Chafe, 1994) is known to be associated with an activation cost, higher for the introduction of new information and lower when processing textually given information and identity relationships. In terms of psychophysiological models of sentence comprehension, the higher is the cost to reach full activation of the currently processed referent, the stronger will be the effort in order to retrieve lexical and semantic information and to integrate it within the currently processed partial utterance. A measure of the activation cost, and of the cognitive effort to integrate the processed element within the available representation of the sentence information can be obtained by observing the modulation of the Event-Related Potentials (ERP), time-locked to the target element. The modulation of specific ERP is in fact known to reflect specific cognitive mechanisms, such as for example, the processing of morpho-syntactic aspects of language (e.g. agreement, word-order violations)(e.g., Steinhauer, Drury, Portner, Walenski, & Ullman, 2010), the integration of semantic representation of words (Federmeier & Kutas, 1999) and the auditory response to sound manipulation (Näätänen, 1990; Näätänen, Kujala, & Winkler, 2011; Näätänen, Paavilainen, Rinne, & Alho, 2007). Currently, there is a growing body of literature addressing also the processing of discourse structure, and the formation of discourse model (Burkhardt, 2006, 2007) as well as the role of prosody in the decoding of spoken language (Dimitrova, Redeker, Egg, & Hoeks, 2008; Heim & Alter, 2006; Magne et al., 2005; Schumacher & Baumann, 2010; Steinhauer, 2003b).

4.3. Psychophysiological evidence about the integration of Intonation and accessibility of referents

I present, here, some specific ERP components that have been found in relation to both integration costs and the discourse updating processes, and specifically the manipulation and results of Schumacher & Baumann (2010).

Early ERP experiments investigating bridging inference were performed in the written domain.

Burkhardt (2006, 2007) respectively compared the processing of identity and bridging relationships and the processing of inferential bridging of different types and found that the activation cost of a referent elicited two main ERP components: a negative peak with a latency of about 400ms (N400) and a positive shift at about 600 ms (late positivity).

Positive deflections starting at about 600 ms elicited by semantic and pragmatic manipulation are typically generically referred to as Late Positivities and have been found to reflect the computational costs of anchoring in anaphoric relationships (Kaan and Swaab, 2003; Burkhardt, 2005).

The N400 component has been initially linked to the detection of contextual incoherence and semantic violations (e.g., Federmeier & Kutas, 1999; Kutas & Hillyard, 1984; Kutas, Lindamood, & Hillyard, 1984). Following studies have showed that its amplitude becomes larger as the semantic plausibility or expectancy of a word or concept lowers (in terms of cloze probability) in a given linguistic or non-linguistic (e.g., a pair of pictures matching) context (e.g. Hagoort, 2005; Hagoort & Brown, 1999; Kutas & Federmeier, 2000; Van Petten & Kutas, 1991).

Currently, the interpretation of the N400 component is debated between two opposite and polarized perspectives.

On one side, the N400 component is interpreted as reflecting the incremental semantic processing of the “activated word meaning with sentence context and general world knowledge” (Nieuwland et al., 2019; p. 3 of the preprint version). Within this view, the processing is intended as a compositional integration of the meaning, from the word-level, and facilitated by the predictability given by the world knowledge (Van Berkum, Hagoort, & Brown, 1999; Brown & Hagoort, 1993). In this perspective, the incoming word is immediately related to the semantic representation of the previous contextual information (Van Berkum et al., 1999).

On the other side, the N400 component is interpreted as reflecting the access to meaning. Thus, the preceding context is assumed to activate a set of related meanings, and within this activation also occurs the preactivation of the Target word. Within this view, the integration of the word meaning with the context does not occur until 400 ms have passed (Bornkessel-Schlesewsky & Schlesewsky, 2008; Kutas & Federmeier, 2000; Nieuwland et al., 2019).

Recently the debate has concentrated on this debate and began to propose a unified perspective that indeed supports the presence of both mechanisms. The ability to predict a target word within a specific context and in the presence of pre-activation has been found in several studies (Van Petten & Luka, 2012). On the other side, a significant amount of evidence is available to support the view of an N400 modulation in function of semantic plausibility (Li, Hagoort, & Yang, 2008; Steinhauer, Royle, Drury, & Fromont, 2017).

The evidence supporting both perspectives has led to the formulation of the “multi-process” approach (Baggio & Hagoort, 2011). This view predicts that the N400 reflects the sequential process of first predicting the upcoming word, and sub-sequentially integrate it.

In a recent study, the effect of both semantic plausibility and contextual activation predictability have been tested by Nieuwland et al. (2019). They found that more predictable nouns elicited smaller N400 responses in a time-window between 200 and 500 ms. Plausibility, on the other hand, showed a statistically significant effect only later (650 ms), and with a shallower modulation. They conclude supporting the presence of both mechanisms. Predictability is reflected in the presence of a negativity at 200 ms. The effect of plausibility, on the other hand, is reflected in a smaller modulation at about 350 ms.

The presence of both mechanisms within the N400 modulation is in line with what we have previously discussed in the introduction of the previous chapters and sections. Namely, that the presence of a linguistic context can activate, through a top-down linguistic mechanism the upcoming word. Nonetheless, the integration of the meaning on the basis of the signal has to occur also for never mentioned unrelated items.

Within this perspective, the N400 is intended as “hybrid” component integrating both “compositional and non-compositional processes” (Nieuwland et al., 2019; p. 17 – from a preprint version).

Specifically pertaining the discourse level, Burkhardt (2006) found a correlation between the N400 amplitude’s modulation and the degree of contextual saliency of a referent – the more a referent was contextual salient and identifiable, the lower was the amplitude of the N400 component (see Chapter 1, paragraph 1.2, for a description of contextual saliency). Moreover, she found that the integration of a new independent referent, previously never mentioned in discourse, correlated with the amplitude's modulation of the late positivity component – the newer the information, the greater the positivity. An analogous effect was found by Burkhardt (2007) that has shown that the complexity of an inference that is needed to link a referent within the context increased the amplitude of the late positivity component, with no impact on the N400 amplitude.

Finally, Schumacher and Baumann (2010) investigated the on-line processing of different prosodic realizations of accessible information in whole-part inferable relationships, by measuring ERP while subjects listened to auditory presented pairs of sentences as in (1). They measured ERP during the processing of the target referents (die Sohle, underlined in 1) in sentence-final position in the second sentence. Target referents always constituted textually accessible information because they were in a whole-part relation with the whole referent that was mentioned at the end of the first sentence (Schuh, in 1).

1. Sabine repariert einen alten Schuh. Dabei zerschneidet sie die Sohle.
   'Sabine repairs an old shoe. In doing so, she cuts the sole.'

The second sentence was uttered with 3 possible PAs: H*, H+L*, Deaccentuation (low pitch contour). On the basis of Baumann and Grice (2006), H+L* was assumed to be the preferential prosodic marking for this specific accessible status of a referent, H* was assumed to preferentially align to completely new independent referents and Deaccentuation to be the preferred Intonation for identity relationships, strictly given information. H+L* represented, thus, the experimental congruent condition, and H* and Deaccentuation, the two incongruent, violating conditions.

The results of Schumacher & Baumann highlighted a 3-way modulation of the N400 component: the N400 was, in fact, smaller for the descending PA, H+L*, that is preferentially associated with accessible, semi-active information; an intermediate-level larger amplitude of the N400 was found for, and high peak PA, and a much larger amplitude of the N400 was found for Deaccentuation. This can be interpreted assuming that the correct PA favors the ease of bridging inference needed to integrate the referent within the context. With respect to the other two conditions their data suggested that in a whole-part relation the integration is much harder when the Intonation signals Given information then when is signals New information, in line with previous experiments in the visual domain, and Baumann & Grice (2006)'s findings. Deaccentuation, as a reflection of a highly activated referent, is not appropriate when a sub-specification of a superordinate concept is introduced with a new, previously never mentioned, referent. This is also congruently observable in the following positivity, where they found that the Deaccenting condition, signaling Given information, showed the larger effect, with respect to the other two conditions that did not significantly differ one to the other. The absence of a significant difference in the late-positivity time window, between the processing of H* (signaling completely new information) and H+L*(signaling accessible information), has been explained with the absence of conflicting information in the two PAs. The fact that a newly introduced referent was both yet linguistically inactive, but referentially already semi-activated was reflected in an equal cost in the update of discourse model reflected by the late-positivity component.

The congruent connection between the N400 and Late-Positivity effect called, thus, for the emerging of a biphasic pattern reflecting the role of prosody in the construction of the mental representation of discourse in languages displaying the association between prosody and degrees of activation of a referent, like German (and probably other West Germanic languages).

At different processing stages, prosody showed to integrate the referential accessibility and the implementing of further congruent information in the model. The newly introduced referent can be anaphorically, explicitly or implicitly, linked to the available active information in discourse.

4.4. Italian L2 learners of German

As we have seen in Chapter 1, paragraph 1.2.7., a debated assumed typological difference between West Germanic and Romance languages is the possibility in the first group of languages to modify the accentual pattern in function of Narrow Focus elements, keeping unvaried the syntactical organization of the sentence (Avesani et al., 2015; Swerts et al., 2002). This structural possibility leads to the presence of Deaccentuation in function of the pragmatic structure, independently from the metrical position of the accented syllable, overcoming the phonological rule associating an NPA to the metrical head of the utterance. The phenomenon has been originally labeled with the term: plasticity (Vallduví, 1991).

In Chapter 1, I have extensively addressed the source of confusion generated by evidence coming from experiments designed either to investigate the prosodic marking of referential Givenness (Referential approach), thus investigating the presence of Deaccentuation aligned to textually Given referents, or the shift in earlier positions in the utterance of accentual prominence, normally aligned with the metrical head, caused by the presence of Narrow Contrastive Focuses. The latter case, generating a Deaccentuation pattern aligned to background information within a Focus-Background description (Relational approach).

The investigation of the perceptual processing of a semi-active referent in sentence-final position in German, shifting from being prosodically fully marked, to lacking any prosodic movement – Deaccentuation – represents an interest phenomenon to investigate in L2 learners that share an L1, in which Deaccentuation in sentence-final position represents, at least a dispreferred accentual strategy (or adopting a very strict notion of plasticity, in languages in which Deaccentuation overcomes the phonological rule assigning prominence to the nuclear position). This is the case of Italian, as described in Chapter 1, paragraph X.

Italian, as described, allows to prosodically mark different types of focusses (e.g., Broad vs. Narrow, Informative vs. Contrastive) through the use of different PAs (Gili Fivela et al., 2015; Face & D'Imperio, 2005). The presence of different varieties does not allow to fully generalize the phonological inventory to the language spoken independently from the region of origin, but I consider in this thesis the evidence reported by Gili Fivela et al. (2015) recognizing H+L* as the non-marked prosodic marking of Broad Focus statements, L+H* as the marking of Narrow Contrastive Focus, and H* and L+H* as shared accentuation patterns to mark Narrow Informative Focus.
Nonetheless, in Italian, much less is known in terms of prosodic marking of the referential status of a referent, but the growing body of literature, as described in Chapter 1 paragraph 1.2, consistently reported the lack of use of Deaccentuation, in particular in sentence-final position.

A particularly relevant fact for this thesis, is the fact that it has not yet been fully and clearly explained which of the two mechanisms – Deaccentuation in function of contextual saliency vs. Deaccentuation in function of the presence of shared information and introduction of a new focus element earlier in the sentence – guides the resistance to the use of Deaccentuation. The lack of the use of Deaccentuation, as presented, is particularly evident in studies investigating the L2 acquisition of phonological and phonetic aspects of West Germanic languages, in adult L2 learners, native speakers of Romance languages, and specifically native speakers of Italian (Avesani et al., 2015; Cruttenden, 2006; Rasier, Caspers, & van Heuven, 2010; Rasier & Hiligsmann, 2007).

The second aspect that motivated the choice of native language speakers of Italian, is the fact that in this particular experimental manipulation, both in German and Italian, the congruent experimental condition the falling contour H+L*: in German, it is congruent as prosodic marking of a semi-active referent, in Italian, it is congruent when aligned to non-marked Broad Focus statements. Since the sentence-pairs adopted by the originally study of Schumacher & Baumann are indeed constructed as a couple of Broad Focus sentences, in which the decoding of the context sentence does not lead to the expectation of any specific Narrow Focus elements, the speakers of the two languages in their native system can find H+L* as congruent manipulation. This leads to assume that in case of transfer strategies in the perception of stimuli uttered in L2 German, the L1, in this case, can favor the experimental manipulation helping the recognition of the appropriate pitch manipulation also in L2. This possibility does not represent a bias condition since the congruent condition is, in fact, the control condition of the experiment.

The primary focus of this experiment is thus to investigate how the information status of the target referent, representing semi-active information, expressed in a non-marked Broad Focus structure, and congruently marked with H+L* for both native speakers of German and Italian, is processed in L2. In particular, the aim is to observe whether the graded modulation of accessability of the referent associated to 3 specific pitch contours, is reflected in a graded modulation of the N400 component, followed by a positivity larger for the Deaccentuation condition, as in native speakers of German (Schumacher & Baumann, 2010).

The second explicit focus of this experiment to investigate by means of the EEG technique the online processing of Deaccentuation, to shed more light on the perception of the lack of prosodic marking in nuclear position in L2, which can provide relevant information for the L2 acquisition of the intonational systems of West Germanic languages, by native speakers of Romance languages.

In fact, what is shared across the presented acquisitional studies of prosodic marking of referential status in L2 is the fact that the modulation of pitch in function of the pragmatic organization of discourse represents a marked feature to acquire, compared to the prosodic in function of the metrical structure only (Avesani et al., 2015; Swerts et al. 2002; Rasier et al. 2010). This has been particularly highlighted by studies requiring to re-construct the information structure of the discourse on the basis of the auditive perception of prominence patterns both in cross-linguistic comparison of West-Germanic and Romance languages (Swerts et al. 2002) as well as in SLA studies concerning languages with the same typological differences (Rasier et al. 2010). These studies highlighted that native speakers of Romance languages fail in reconstructing the information status of a referent on the basis of the perceived prominence and that the proportion of accurate reconstruction is significantly lower in L2 than in L1.

A different perceptual strategy of the Deaccentuation contour can, on the one hand, be motivated by the presence of sensory mechanisms at a lower cognitive level, as the ones presented in Chapter 2 and 3. In this case, the processing of Deaccentuation in L2, in native speakers of a language that does not make systematic use of it, can be influenced by phenomena such as "apperception" (Gass, 2017) modulating mechanisms of selective attention towards to the relevant pitch modulations of the native language. Deaccentuation is this case could be perceived as phonologically not relevant, given the modulation present in L1 and the fact that is acoustically less salient than the normal modulation expected in nuclear position. Therefore, the resistance found in the L2 acquisitional process of native speakers of Italian and Romance languages could be reconducted to a substantially different low sensory perceptual strategy. If attentional mechanisms are the basis of a different processing of Deaccentuation, a possible acquisitional training strategy would be then the instantiation of the process of category formation, opposed to a mechanism seeking a recognizable category (see Chapter 1 and 2), in which the pitch modulation corresponding to Deaccentuation, a low flat contour, will have to be systematically mapped and recognized in the presence of referentially given information. This auditive response to the perception of pitch movements is observable in the modulation of early ERP components (before 300 ms); nonetheless, it was not previously and specifically addressed in the study of Schumacher and Baumann (2010). The investigation of this aspect represents; thus, an improve of the paradigm originally used.

On the other hand, the processing differences in L1 and L2, given the multidimensional nature of intonation, could involve higher-level cognitive processes, involving the integration of the semantic information at word and discourse level, as well the processing of the syntactic structure.
In this case, the hypothesis on the processing of incongruent prosodic conditions is two-fold.
On the one hand, the way bridging inferences are processed and semantically solved should be shared by speakers independently from their native language, reflecting a universal cognitive process.
If intonation universally is an instrument to acoustically guide the hearer towards the relevant and newsworthy part of information conveyed in the discourse (Pierrehumbert & Hirschberg, 1990; Gussenhoven, 2002), then both language-groups should show a sensibility, at least two-way, between what they consider the appropriate and congruent pitch modulation in sentence-final position of an accessible referent. For the speakers of both languages, the target NP represents a new linguistic term introduced in the model, partially already activated by the decoding of the first context sentence, in which the whole referent is present. Therefore, it is expectable to find a similarity across language-groups in the processing of the congruent condition (H+L*), also favored by the overlapping Broad Focus structure.

On the other hand, linguo-specific properties showed to affect the perception and acquisition of marked relationships between intonation and information structures in L2, and native speakers of Italian have been repeatedly found to fail to Deaccent referential Given Information (Avesani & Vayra, 2004; Avesani et al., 2015), and to modulate the association between intonation and information structure on the basis of relational properties linked the prosodic realization of focus (Face & D'Imperio, 2005) and to structural metrical properties of the phonological system (Avesani 1995, 1997). Thus, processing differences of the incongruent conditions, especially of the Deaccentuation conditions are expected to be observable in the modulation of the biphasic pattern reported by the studies of Burkhardt (2006, 2007) and Schumacher and Baumann (2010), Baumann & Schumacher (2012).

4.4.1. Specific predictions of the processing of the Deaccentuation condition in L2

On the basis of the presented literature, it is reasonable to expect when observing the matching and mismatching prosodic marking of Referential Semi-Active information in L1 and L2 (Italian L1), that the following aspects emerge:

1) The properties of the intonational system of Italian, by default, not making an equivalent use of Deaccentuation to the same extent as German in West-Germanic languages should, in first instance, affect the auditory processing of matching or mismatching contours.
   In particular, if the native language warps the perceptual space of L2 learners (Kuhl, 1991) and affects the auditory perception in L2 filtering out non-available sound representations or converging L2 diverging phonetic properties to the available L1 sound representation (Flege, 1995; Best & Tyler, 2007; Mennen, 2015) then ERP patterns should show that Deaccentuation is differently evaluated by L2 speakers, compared to L1, and possibly not considered by auditory attentional mechanisms.

2) If intonation is a crucial aspect in the construction of the mental representation of discourse, then a different evaluation of Deaccentuation must emerge in ERP components found to related to the discourse processing. In particular Deaccentuation if not analogously acoustically processed by L2 speakers should also not be a reason to impair the integration of a referent in the currently activated representation of the discourse and therefore elicit a lower amplitude negativity at 400 ms in respect to native speakers, and possibly not displaying at all an N400 effect in respect to the appropriate prosodic marking of accessible information.

3) For analogous reasons, the processing in L1 and L2 should imply a different updating cost in the later time window. This is motivated by the fact that the late-positivity appeared to be sensitive to the introduction of a new referent, despite the referent is semi-activated by the bridging inference, and that auditorily this is reflected in a higher cost when the prosodic marking signals through Deaccentuation that the currently processed referent is actually already active, when it is not. Again, if the L1 system modulates the perception of Deaccentuation, Deaccentuation should not be in L2 a source of higher cost.

4.5. Differences in the implementation of the present study vs. Schumacher & Baumann (2010): A Contrastive Focus accent.

A relevant aspect that could have affected the L2 processing of Discourse Structure and biased the results, is the relative higher difficulty of L2 learners to segment natural speech produced by native speakers during auditory perception (Field, 2003; Finn & Kam, 2008).
In a second language, it is, in fact, harder to correctly segment the continuous flow of speech, recognize and access the meaning of every single word in the utterance.
The difficulty in the perceptual task of segmentation can be increased both by the presence of low-frequency or highly specialized vocabulary, requiring to disentangle words that are rarely used by L2 learners, and by the rate of speech of the speaker, especially when visual cues like labial movements are missing (e.g., telephone conversations) during the perception task.
Firstly, the nature of the experiment requiring to perceive a long battery of sentence-pairs in German L2 without visual information on the vocal tract, made necessary to adopt a choice of vocabulary suitable for both the needs: the necessity of a high number of stimuli, and to overall keep a frequently used vocabulary.
In second instance, since the original experiment was designed for native speakers, the newly designed sentences have been re-recorded with a slower speech-tempo and a modulation of the speech style

adopted in the Second Language Teaching context. The chosen speaker was, in fact, a professional L2 instructor for German, specialized in phonetic training. This resulted in a structural change of the H* PA in L+H*, caused by the fact that the pitch contour signaling a high tone has been systematically produced, across all stimuli, with an extremely high peak and very steep F0 rising. The degree of systematicity with which the professional speaker produced this specific pitch contour led to the decision of not further manipulate this peak to achieve the intended H* accent.

This has changed the initial idea to compare only referential aspects of Discourse between two languages that systematically showed a different sensibility to the prosodic marking of the referential status of a referent and led to the possible introduction of a relational modulation such as the marking of a Contrastive Focus (CF).
L+H* is reported in the presence of CF information in both languages: in German … in Italian.

It must be noticed that, the switch from the original experimental design, investigating only the manipulation of pitch movements in function of the cognitive activation of a referent, to the inclusion of a possible relational manipulation caused by a simple variation in pitch height, can be something that regularly occurs in the experimental manipulation of the stimuli in the construction phase of an experiment, therefore it is important to address the role of phonetic implementation having an impact on the phonological representation of an accent. A non-intended variation, like the production of a higher peak, can be a source of the confusion in experimental evidence merging referential and relational approaches since a small change in the phonetic parameters then results in a structural change of the phonological categories.
For these reasons, I decided to make use of the unexpected variation to investigate whether this specific newly introduced modulation impacts first in L1 processing, comparing the ERP elicitation in the paradigm of the original study with the data of the present one. In second instance, since in truth the modification of the paradigm actually perfectly addresses the debate on the possible differential nature of the processing of referential vs. relation properties of the integration of prosody and discourse structure, the introduction of a condition of a different nature is a good opportunity to observe with ERP whether the two violating conditions – the Deaccentuation contour and the L+H* peak (now majorly diverging also in function of acoustical salience - are differently processed at low and high cognitive level in L1 and L2.
For these reasons, I introduce the available evidence on the processing of focus elements in neurophysiological research.

4.6. *Focus-related ERP components*

Psycholinguistic literature on focus-processing reports in several studies the facilitatory effect that focus has in comprehension tasks. Firstly, it has been shown that focused information is generally processed faster than non-focused content (Birch & Rayner, 2010; Morris & Folk, 1998). In second instance, the presence of a focus unit showed to affect the attentional resources in comprehension tasks, allowing to detect violations and incongruences better. The recognition of anomalous words, target phonemes, and single terms varying from one sentence to another showed to improve when the target item was placed within a focused unit generated, for example, by an "it-cleft" structure or guided by a QUD opening a narrow empty slot of requested information (Cutler & Fodor, 1979; Sturt, Sanford, Stewart, & Dawydiak, 2004; P. Ward & Sturt, 2007; for an interesting review on attentional resources' allocation guided by focused structures see (L. Wang, Li, & Yang, 2014; L. Wang et al., 2014). The ERP literature addressing the perception of focused units is wide and complex since the several available experiments found different (combination of) components in function of:

i) the visual or auditive modality;
ii) the type of experimental manipulation, generating predictable focus-background structures by means of the QUD or non-predictable ones;
iii) the explicit or implicit judgment of the prosodic structures required by the experimental task;
iv) the position of the focused unit in sentence-initial, medial, or final position within the target utterance;
v) the violation of the association between IS and intonation by means of the positioning of an unexpected superfluous accent on a background (superfluous accent) partition of the discourse, or the complementary manipulation, that deaccents a predicted focused unit, creating a condition where an expected accent is missing (missing accent).
vi) The interaction of the focus effects with the hierarchical organization of referential givenness, similarly interacting as in L2 literature.

Following the previously mentioned studies, some studies reported that the processing, in medial position of focused words in the visual domain, elicited a smaller N400 than the processing of non-focused words (Wang, Bastiaansen, Yang, & Hagoort, 2011), considering the N400 an indicator of processing difficulty. On the other hand, several studies also in the visual domain, reported that the processing of words in focus position elicited a larger positivity compared to the processing of words in the background position ((Bornkessel, Schlesewsky, & Friederici, 2003 - initial, medial, final position; Cowles, Kluender, Kutas, & Polinsky, 2007- medial and final position; Stolterfoht, Friederici, Alter, & Steube, 2007- medial and final position). The positivity component P3b elicited by focused elements has been interpreted as a sign of integration effort; therefore, the latter studies indicate that focused words require more cognitive resources to be decoded, compared to non-focused terms.
Remaining in the visual domain, Chen, Wang, & Yang (2014) investigated processing differences separately imputable either only to focus manipulation or to the degree of cognitive activation linked to referential Givenness. They created short stories, composed of 3 sentences, projecting on the screen one word at the time. The factor focus could appear with 2 levels (+Focus, -Focus), the factor "cognitive activation," to which they referred as Newness, could also assume 2 levels (New, Given). Each story then could appear in 4 versions as reported below (the indication of appropriateness is mine):

1) Heren (masculine) was persuading his friends to go on an outing. He ignored that the weather forecast had predicted bad weather. At the time **Zhongying** (feminine) [-focus, new] reasonably opposed him. (appropriate)
2) Heren was persuading his friends to go on an outing. He ignored that the weather forecast had predicted bad weather. At the time, it was **Zhongying** [+focus, new] (who) opposed him reasonably. (informative narrow focus appropriate, contrastive narrow focus inappropriate)
3) Heren was persuading Zhongying and others to go on an outing. He ignored that the weather forecast had predicted bad weather. At the time **Zhongying** (feminine) [-focus, given] reasonably opposed him. (appropriate)
4) Heren was persuading Zhongying and others to go on an outing. He ignored that the weather forecast had predicted bad weather. At the time, it was **Zhongying** [+focus, given] (who) opposed him reasonably. (Narrow and Contrastive focus appropriate)

They analyzed the ERP components in 3 main time windows: 150-250 ms, 250-500 ms, 500-700 ms. In the early time-window, they found a main effect of Focus in anterior regions, more robust in the Left Anterior quarter. Focused words elicited a larger positivity compared to non-focused words, and new information did not elicit any significant effect. They interpreted the positivity as a P200 component and functionally explained it as an indicator of attention allocation (Carretié, Mercado, Tapia, & Hinojosa, 2001). In the second time-window, they found an opposite effect. Focused words elicited a smaller N400 in frontal electrodes compared to non-focused words; words carrying new information elicited a larger negativity, more robustly in the central region of the scalp, compared to referentially given referents. No interaction between Focus and Newness was found.
Chen and coauthors' functional interpretation of the smaller negativity for focused word, which is possible to interpret also as Late Positivity Component (LPC), is in line with the one provided in the previous studies (Bornkessel et al., 2003) and supports the view that readers engage in reading focused words to a deeper extent compared to when they engage in reading non-focused words. Chen and colleagues also considered possible alternative explanations that supported the view that the presence of the P2, granting more attentional resources, can cause a facilitator effect in the integration process. They also considered the possibility of interpreting the positive shift as CPS (closure-positive shift), given the fact that they experimentally induced the focusing effect by using a sentence structure corresponding to the it-cleft structure in Chinese. The use of this specific structure could have caused the readers to mentally place a pause within the utterance, and it could have created a phrasing effect.
The smaller effect of the N400, found by Chen and colleagues in correspondence of given elements compared to new elements, could be both interpretable as a sign of higher integration costs required by new information, or also as an effect of priming since the referents were textually given.
In the last time window, focused words elicited an enhanced broadly distributed late-positivity component (LPC) and a reduced anterior LPC in the presence of new information compared to given information. No interaction between the focus and newness factor was found.
The presented experiment is very interesting since it allows to disambiguate the effects of the relational approach in opposition to the referential perspective and it strongly signals that the two factors don’t interact.
Nonetheless, the authors observed only the comparisons within the single referring factor (focus, or newness). Even though they introduced completely new or completely given information in the discourse,

they did not consider any factor of appropriateness of the specific focus manipulation, but on the contrary, they have considered every condition as congruent.
In the auditory domain, on the contrary, a sentence like in the example number 2 in the stimuli, depending on the PA aligned to the target referent Zhongying, can introduce either a narrow informative focus or a contrastive focus. If the new referent Zhongying were introduced by means of a narrow informative focus, implicitly signaling also completely new information, the sentence would have been considered marked, but still congruent to the context. On the other hand, a sentence, in which a contrastive focus was aligned to the referent Zhongying, but the referent was not either previously mentioned, nor inferably anchorable to another identifiable term (to which it could be contrasted to), the manipulation would have been infelicitous. This would have happened because the accent would have signaled to search for an anchor in the previous context, and the anchor could not have been found. This paradigm, therefore, is very interesting, nonetheless remaining in the visual domain does not consider the information conveyed by the prosodic information.

In the auditory domain, two main types of paradigm have been adopted: one using a Question Under Discussion (QUD) (Büring 2007) manipulation, thus inducing an expectation on the presence of Narrow Focus; and one using sequences of sentences pragmatically and semantically coherent between each other, marked by congruent or incongruent prosodic information (as the paradigm adopted by Schumacher & Baumann, 2010). In both paradigms, it can either be required to participants to specifically evaluate the prosodic contour (prosodic judgment task) or to answer questions on the basis of the stimuli (comprehension task). The pitch manipulation can be in sentence-initial, medial or final position.

In two studies (Hruska & Alter, 2004; Hruska, Alter, Steinhauer, & Steube, 2001), adopting both the a QUD induced expectation and the prosodic judgement paradigm, it has been found that a missing PA on focused elements elicited respectively a N400 component and a biphasic N400-P600 pattern, while a superfluous accent on background information did not elicit any effect. Moreover, the congruent and expected presence of a PA aligned to new information elicited an Expectancy Negativity (EN) in frontal electrodes, and posteriorly a CPS in a later time window.

Adopting an analogous paradigm, Ito & Garnsey (2004), found in sentence-initial position a posterior positivity between 250-500 ms for missing accents on focused elements, and a frontotemporal negativity in sentence medial position. No effect was found for superfluous accents. Also, in an analogous paradigm, but adopting a contrastive focus marking on given information, Magne et al. (2005) found, for both missing and superfluous accents, in medial position a P600 effect and in sentence-final position an N400 effect.

Summarizing the main results of these studies, the prosodic judgment guided by the expectation generated by a previously perceived context question, led to the elicitation of an N400 effect, signaling a higher integration cost, when the expected PA aligned to focused elements was not encountered. In few cases, the missing accents also elicited a late P600 effect. Superfluous accents overall did not elicit any specific component, apart from the results of the experiment of Magne et al. (2005), in which incongruous contrastive focus accents in sentence-final position elicited an N400 effect.

In studies adopting the QUD manipulation in the context sentence, and a comprehension task (Dimitrova, Stowe, Redeker, & Hoeks, 2012, p. 2; Dimitrova et al., 2008; Toepel, Pannekamp, & Alter, 2007): Toepel et al., 2007 and Li et al.,2008 found an N400 effect for missing accents and no effect for superfluous accents.

The congruent focus-accent manipulation of new information elicited a late positivity interpreted as CPS; on the contrary, Dimitrova et al., (2012) found a P600 effect for missing accent, and a 3-phasic modulation for superfluous accents composed of: an early positivity (100 ms), and an N400 effect, followed by a P600 effect. Also, Dimitrova and colleagues did not report the late positivity effect found for the association between focus and accentuation.
As Li et al., (2017) reminded on the basis of the results of Chen (2014), the studies of Hruska and colleagues, and Toepel and coauthors associated the focus position with new information, while Magne et al., (2005), Cowles et al. (2009) and Dimitrova et al. (2012) aligned and/or created a contrastive focus structure associated to given information.

It is possible therefore to hypothesize that, the congruent association of focus marking and new information in QUD paradigms generates the repeatedly reported effect of a late-positivity and that, within this type of manipulations, an N400 effect is reported for missing accents and no effect is reported for superfluous accents. Moreover, in experiments, in which the attention of the participants is not guided to specifically attend a precise pitch accent in a specific position, but that on the contrary, require to decode the prosodic information to successfully comprehend the stimuli, the encountering of an incongruent contrastive focus marking in sentence medial position (superfluous and missing accent) elicits a P600 effect, and superfluous accents are detected early in two time-windows: before 200 ms and before 500 ms.

4.7. L1 vs. L2 processing of focused referents (interacting with referential givenness)

This study in final instance aims to compare L1 results with the original Schumacher & Baumann (2010) results in order to highlight processing difference between H* and L+H* given the same information structure of the sentences. Differences imputable to the introduction of a narrow-focused unit within the same experimental paradigm could add evidence in line with the direction opened by Chen et al., (2014) and Li et al., (2017) highlighting the necessity to investigate the processing of these two phenomena separately. Since this was not the original aim of this study, the manipulation adopted here will not consent to discuss the results excluding an interaction between the two factors, but it will exploratively report whether the resolution of a bridging inference, targeting the part referent, is affected by the shifting from a broad focus informative prosodic manipulation to a narrow contrastive one, and namely whether the focus manipulation elicits components in accordance with the results of Magne et al., 2005 and Dimitrova et al., 2012, for superfluous accents, even though, here a question-answer paradigm is not adopted and the manipulation occurs in sentence-final position.

In second instance, in Italian, as previously mentioned, the prosodic marking of contrastive information can vary across varieties but in several cases, either H* or L+H* are reported as common PAs for narrow contrastive focuses (for a comprehensive review of intonational differences across varieties see Gili Fivela et al.,2015). Therefore, the change of PA from H* to L+H* in principle should not affect the predictions on L2 learners in this condition, since also within the original paradigm there could have been the possibility the H* in sentence-final position could have been interpreted either as narrow informative or narrow contrastive focus. The difference between these two accents is by definition controversial but mostly defined by a phonetic difference of the F0 contour. What mostly allows to identify the L+H* accent correctly is the steep F0 rise and the expanded pitch range (Grice et al., 1996). This is exactly what characterize the stimuli produced by the speaker in this condition.

Therefore, in an L2 perspective what indeed this new accent introduces is a highly detectable acoustical salience, granted by the prominence, of the final NP. In the original design, it could have been expectable that the modulation based on the degree of givenness, slightly changing the pitch modulation in final position without any other structural difference, might not have been detected by L2 learners because strongly relying on the matching of the BF with a metrical head. In this design, introducing L+H*, that also clearly matches a focus modulation across varieties, it is expectable that the evident acoustical salience could be detected.

The paradigm, following the parameters of the Italian structure, tests against the congruent condition H+L*, the lack of accentuation (missing accent) and the alignment of an overtly salient and structurally not necessary F0 peak, that also instruct the hearer to identify and contrast a previously mentioned anchor (superfluous accent).

4.8. *L2 Learners profile – internal vs. external factors affecting L2 perceptual processing.*

Finally, I briefly introduce the correlated measures that will be analyzed in relation to the modulation of ERP components in L2 processing.
In Chapter 1, paragraph X, I have presented the internal and external factors affecting the potential success of the L2 acquisition of phonological and phonetic aspects in late learners.
External factors are the variables commonly playing a role in L2 acquisition, that can be reconducted to the fundamentals concepts of the field of SLA: namely, they are responsible for the main evidence signaling that the perception and thus acquisition of phonological aspects in L2 is affected by the presence of the phonological categories and phonetic features present in the native language, acting as a "filter" in the processing of the L2 speech signal.
The role of the native language, as reported in the introductory chapter, is found to have a different impact on language learners in function of their age, and in function of the specific L2 feature that has to be acquired.
For what pertains L2 acquisition of phonology, it has been widely reported that the Age of Acquisition (AoA) affects the rate of successful mastering of these features in L2 as well of the Age of Arrival in the country in which the L2 is spoken as standard language. A third factor that externally models the rate of acquisition is whether participants received formal education in a class-settings, or simply arrived in the country in which the L2 is commonly spoken and began the process of acquisition as naïve learners without any instruction.
In this experiment the group of L2 learners is selected meeting a main criterion of representing a population of learners ranging from intermediate to advance levels of acquisition, therefore mainly representing a population of learners spanning from the B2 to C1 European Reference level (see the method section). This means that the sample of participants could vary in terms of Age of Acquisition and years of instruction. In second instance, the experiment investigates specifically the prosodic processing of learners living in a German-speaking country for at least 2 months at the time of the experiment. The criterion allows, therefore, the presence variability in function of the Age of Arrival and length of the period of immersion in a German-speaking country.
Since this experiment is one of the firsts investigating the L2 acquisition of prosody by means of EEG measures, information from participants about relevant variables has been collected, and a specific aim of the study is to investigate the impact of the external factor in the modulation of the elicited ERP components, in function of prosodic manipulation.
I have collected therefore data on: the biographic characteristics of the participants; their general musical education and competence; the type of education pertaining German they had received prior to the experiment within scholastic and university education, and in private classes; the age at which they began to study German and arrived for a prolonged time in a German-speaking country. In final instance, the questionnaire asked about the usual length of exposure on a weekly basis to audio-visual material spoken in German.

Internal factors on the other hand, as described in Chapter 1, are all the variables impacting on the L2 acquisition at the individual level, shifting from the variability present across individuals in different areas of cognition such as memory capacity, modulation of selective attention, and discrimination capabilities.
Internal factors also involve aspects of the cognitive profile and personality that modulate the presence of extraversion vs. introversion, the facility to communicate and interact with other individuals, the presence

of empathic traits and of traits belonging to the construct of Theory of Mind (ToM), allowing to correctly consider and evaluate the behavior of an interlocutor in function of his perspective, that can differ from the one of the speaker.

In Chapter 1, I have introduced, how intonation and the characteristics of the ToM construct are related because of the role they play in efficiently decoding the communicative intentions of a speaker or specularly of a hearer. A speaker can intentionally highlight specific elements of discourse, because assuming the interlocutor will find them coherent, informative, and relevant within the discourse model, affecting, therefore, the manipulation of Focus-Background structures. Besides, I have introduced how specific traits belonging to the Autism Spectrum Conditions (ASC) specifically capture this relation, correlating with the reported evidence of the population with ASC, showing impairments both in tasks requiring the evaluation of pragmatic aspects of language, and the appropriate modulation, and recognition, of intonational contours.

For these reasons, as in the previous experiments, I have adopted here as correlated measure of internal factor playing a role in the L2 processing of pitch movements the scores of two questionnaires: The Autism Spectrum Quotient (AQ) (Baron-Cohen et al, 2001) and the Interpersonal Reactivity Index (IRI) (Davis, 1983, Albiero et al.,2006) (see Chapter 1, paragraph 1.6.4. for an overview on the tests).
The AQ is composed of 5 subscales, each investigating the presence of specific aspects and behavioral patterns known to connotate individuals with ASC. Since two of these subscales respectively investigate the presence of an unusual attention to details (Att_Dett), and the difficulty in switching the attention towards new relevant stimuli (Sp_Att), the use of the AQ as covariate measure allows, in addition to the evaluation of an overall pragmatic profile, to specifically evaluate the modulation of attention of the participants.

Both, the literature of focus marking, performed with EEG measures, and the literature addressing perception within the field SLA highlighted a role of an attentional mechanism, respectively in the processing of focus elements and in the acquisitional phase, in which participants notice the phonetic features of the L2 stimuli – apperception - and evaluate the nature of the stimuli assimilating them or not to the L1 available categories. Therefore, the presence of covariate measures related to attentional mechanisms can be used to investigate their relations with the modulation of specific ERP components. Namely, it is possible to observe if attentional mechanisms are differently impacting the processing of the Deaccentuation condition (missing accent) and of the CF condition (superfluous accent) at early or late stages and whether the sample of participants is homogenous with regard of this specific internal factor.

4.9. Method

4.9.1. Material and Procedure

Following Schumacher & Baumann (2010), 144 sets of stimuli were created, composed by a sequence of two sentences: a context sentence, containing an anchor referent, and a target sentence having a final critical NP made inferentially accessible by the anchor referent in the context sentence. As in the original experiment, between the anchor and the target NP always stood a Whole-Part relationship, as in the example below:

(1) Sabine findet einen alten Schuh. Dann repariert sie die Sohle.
'Sabine finds an old shoe. Then she repairs the sole.'

The syntactic structure of both sentences has been kept constant throughout the whole experiment. In the target sentence, the conjugated verb always compared in second position and the critical NP always represent an object of the verb in sentence-final position. The rhythmic structure has also been kept constant: the critical NP has been constructed with a definite article followed by a bisyllabic noun carrying the word-stress on the first one. The stimuli were read by a trained phonetician and digitally recorded in a soundproof room at the Institute of Phonetics at the University of Cologne. Prosodically, the context sentence has been uttered with a constant Broad Focus intonation in the whole set. The target sentence has been created with 3 different Nuclear Pitch Accents aligned with the critical NP.
The 3 conditions are: H+L* (Congruent condition = Accessible information); L+H* (originally H*, signaling completely new information – Now: potentially signaling Contrastive information); Deaccentuation of the target NP (marked in figures with L), associated with a nuclear L+H* assigned to the preceding verb, signaling completely Given information. The resulting 432 critical trials were distributed across three lists (with 48 trials per prosodic realization each) and interspersed with 168 (84x2) fillers (Adjectives) 396 (132 x3) (scenario items) = 564. The experimental session contained 360 items and was conducted in 9 blocks with pauses in between. During the experimental session in each trial, a white fixation star was presented in the center of a dark grey screen for 1000 ms, and the trial was presented auditorily while the fixation star remained on the screen. Between the two sentences, 750 ms of silence were inserted. 1000 ms after the target sentence ended a Yes-No comprehension question, assessing the participant's attention, visually appeared in white on the screen. The comprehension question could refer to the context or the target sentence and was equally distributed and randomized across lists. The participant answered by clicking two different buttons on a response box. Half of the participants answered Yes by clicking a button with the right hand, half by clicking with the left hand. The response time was restricted to 4000 ms. (incorrect or timed-out responses were excluded from the ERP analyses.) Each session started with a short practice block to familiarize participants with the procedure. The electroencephalogram has been recorded through 24 Ag/AgCl electrodes on an elastic cap (EasyCap). The signal has been amplified with a BrainVision Brain-Amp amplifier and digitalized at a rate of 500 Hz. The reference electrode was placed on the Right mastoid and re-referenced offline to linked mastoids. The ground electrode was placed on AFz. In order to control for eye-movement and exclude eye-artifacts, two electrodes have been placed above and under the right eye, and two electrodes have been placed to catch horizontal movements on the canthus, one for each eye. All electrodes had an impedance below 5 kO, and all data have been offline bandpass-filtered to exclude slow

drifts in the signal. Only trials without artifacts and correct answers to the comprehension questions were included in the data analysis.

4.9.1. Preprocessing

The recorded data have been bandpass-filtered offline (0.3–20 Hz) before automatic and visual inspection. No baseline correction was applied. The epochs were set between -200 and +1400 ms from stimulus onset and the automatic rejection excluded amplitudes above and below 40 mV the average values. The data of the L1 group of participants have been automatically and visually inspected. The data of the L2 group of speakers only undergone automatic rejection.

4.9.2. L1 speakers

The EEG waveform of twenty-four right-handed, monolingual native speakers of German (5 men, between 20 and 25 years of age, mean=22.4 years; and 19 women, between 19 and 30 years of age, mean = 23.5; overall mean = 23.3) was recorded at the University of Cologne (Germany). The participants have signed written informed consent and received a monetary reimbursement for their time and willingness to participate. Four participants had to be discarded from the data analysis because of excessive ocular artifacts.

4.9.3. L2 speakers

The EEG waveform of twenty-five right-handed, native speakers of Italian, and L2 learners of German (6 men, between 21 and 33 years of age, mean = 26.83 and 19 women, between 21 and 33 years of age, mean = 25.53) was recorded at the University of Cologne (Germany). All participants signed written informed consent and received a monetary reimbursement for their time and willingness to participate. All participants were living in Cologne at the time of the experiment. Before the experiment, they had to fill in a sociolinguistic questionnaire investigating their language habits, and information about the study and acquisition of German as L2 (external factors in individual variability's analysis). All answers were self-declared. In addition, the participants provided a certificate of a recognized organization proving a minimum certified B1 level in German as L2 or attesting the achievement of a relevant competence involving German as L2 (e.g., university degree, job qualification).

Within the screening questionnaire, the questions about the number of spoken L2s and dialects, as well the questions pertaining musical knowledge sought to delineate a general profile of the group, and moreover to observe the presence of strongly monolingual or highly trained, in more than one L2, sub-populations; as well as the presence of highly trained musicians. According to the literature pertaining the language talent (Dogil & Reiterer, 2009; Rota & Reiterer, 2009), trained musicians could have an advantage in tasks involving L2 phonological processing.

Pertaining the knowledge of the local dialect (that in Italian is often recognized as a different self-standing language) and the areas of provenience, the results of the questionnaire indicate the following spoken varieties: 11 participants originally came from the north of Italy, 3 from areas in the center, 7 from the south, and 4 from the two main islands. This supports a general profile of the group as non-variety dependent but heterogeneously representing spoken Italian L1. The profile pertaining to dialect's use showed that: only one participant reported "no-competence and use" of any dialect, 6 reported a general competence in perceiving and comprehending the local dialect but no ability to speak it, 8 to be able to produce only very simple and basic sentences; 5 reported a full bilingual-like use of their local dialect.

Across 4 of the frequently studied European languages, participants reported intermediate to high competences in German as well as English, and in addition, they showed to have also basic knowledge of both Spanish and French, as reported below in Figure 4.9.1.

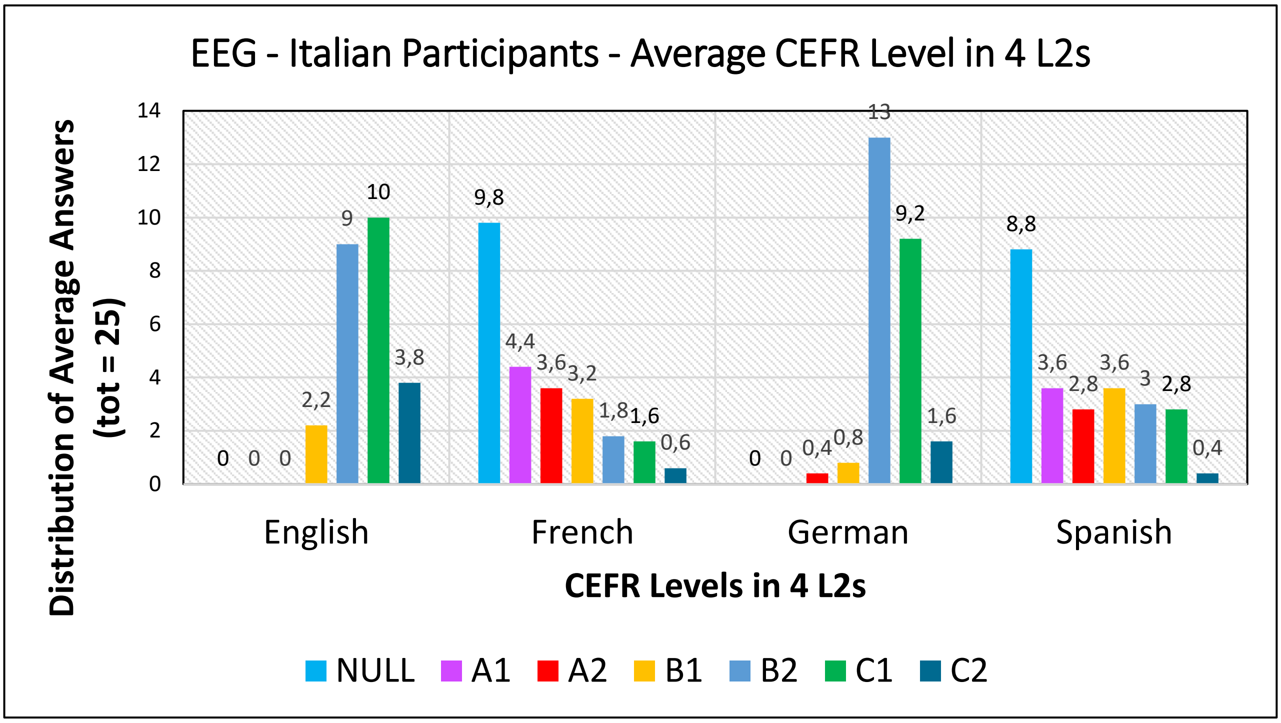


Figure 4.9.1 Distribution of spoken L2

Specifically observing the declared competences in German, participants averagely reported competence in B2 and C1 level in 5 main areas of the CEFR description: Listening, Reading, Interaction, Oral Production, Writing. Only from 1 to 2 participants reported levels either equal to or below the B1 level, and 1 to 2 participants reported the highest C2 level. Distribution of results is plotted below in Figure 4.9.2.:

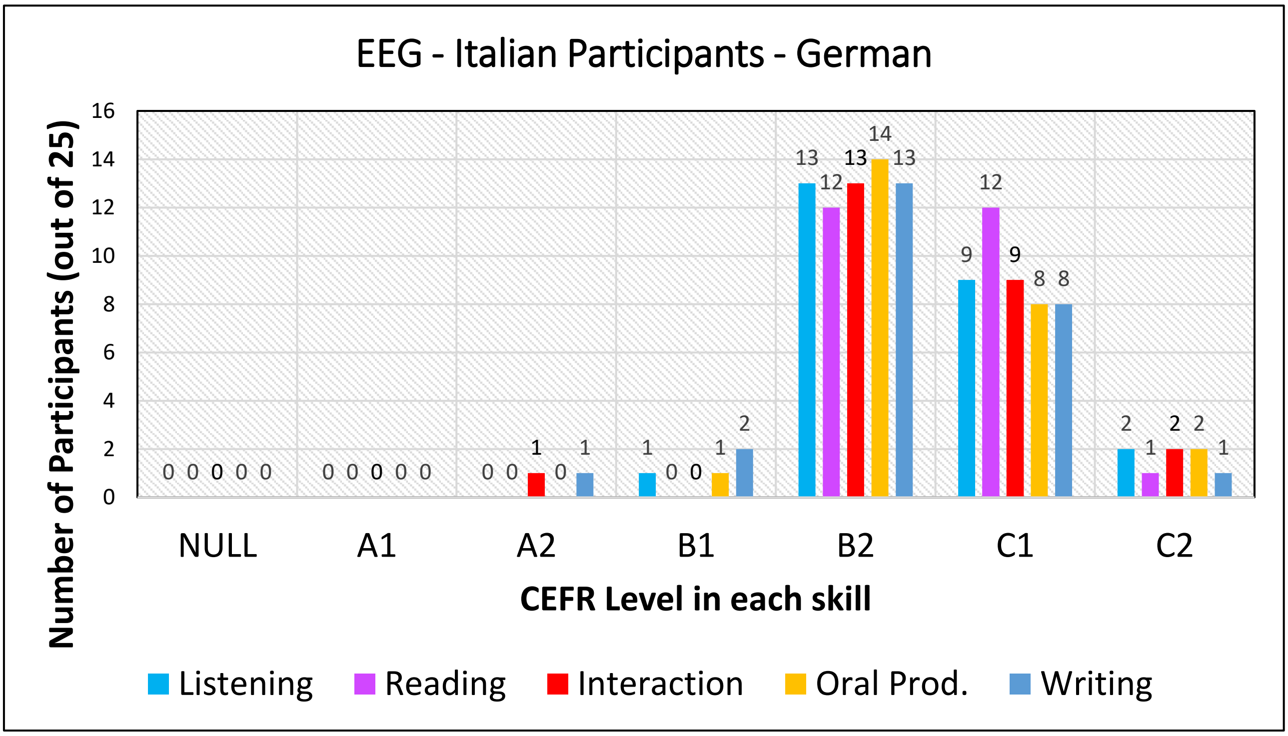


Figure 4.9.2 Competence in German

Pertaining to the musical abilities: only 4 out of 25 participants reported to be actively playing a musical instrument at the time of the experiment. 6 reported from 0 to 3 years of private musical education, and 6 participants from 4 to 10 years of private musical education; 13 reported no musical education at all. In addition, two participants declared 5 to 7 years of training in singing and only 1 declared an official certification of training in theoretical musical skills like solfège. Within the questionnaire, one question asked to correctly identify a passage of a famous composition written for piano on a double pentagram. The task was added in order to identify participants with sufficient musical training to identify and possibly mentally image the melody of the famous composition. The task requires a prolonged an advanced musical education: 3 participants correctly identify the composition. The ones who succeed in the task are 3 participants with a musical education from 5 to 10 years, and one is the one reporting the certification. All 3 reported the task was either rather or extremely hard.

Summarizing, the group appears to be rather homogenous in mastering at least two foreign languages at an intermediate level and showing a general broader competence in more than 2 languages. On the other hand, musical training could be a relevant covariate since half of the population has no training at all, half report at least from 0 to 3 years of training, 3 participants show to be highly trained as musicians, and 2 to be highly trained as singers.

4.9.3.1. Training in German as L2

Out of 25 participants, 14 did not attend any class of German as an L2 during the regular school years (from 6 years of age to 19 years of age). In the remaining 11, only 2 declared to have begun to study German in elementary school, and overall 3 declared from 0 to 3 years of study; 5 declared from 4 to 5 years, and 3 from 6 to 8 years of study.

Data about general education highlighted that only 1 participant, at the time of the experiment, did not hold, or studied in order to hold, a Bachelor's Degree, therefore 24 had a college education: 18 of them were obtaining or had obtained a degree in Italy, 6 of them were obtaining or had obtained a degree in Germany, and they were considered as consistently training the acquisition of German even if they were not specializing in a linguistic education. Out of the 18 participants not completing a degree with German as the language of use, 8 of them were specializing as linguistic mediators, translators or interpreters. Therefore, the total amount of participants specializing in the acquisition of German during the college education was 14. The remaining 10 participants had specialized in another area and learned German subsequently most probably due to the working environment and after having moved to Germany. 2 of the 10 participants without a specific training at university level, were the ones with 8 years of training in German as L2 during school years. Also, the same 10 participants reported from 6 months up to 3 years of attendance in private classes of German as L2.

Summarizing the sample could be possibly further subdivided in a group who began the learning of L2 earlier in life during school education, and participants who acquired it later in life in college or after college education. On the other hand, it is also possible to further sub-divide the sample in participants who systematically trained in German L2 during the years of college education and participants who have specialized in something else and aside have learned German for personal needs.

Looking at the characteristic of linguistic immersion, participants were asked to state the age at which for the first time they continuously lived for at least a month in a German-speaking community (Age of 1st Immersion).

The age ranges from 16 years of age to 32, mean = 23 years of age. The distribution of the Age of 1st Immersion is plotted in Figure 4.9.3 below:

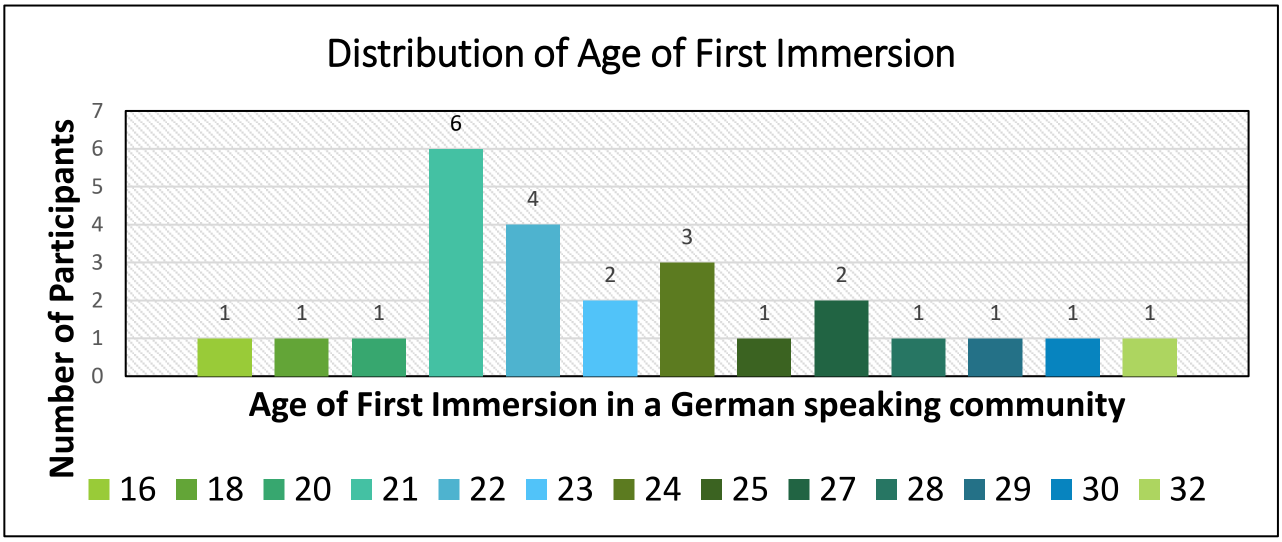


Figure 4.9.3 Distribution of Age of First Immersion

At the time of the experiment, participants ranged from being living in Germany from 2 to 6 months, up from 4 to 7 years of permanence. As the distribution highlights (Figure 4.9.4), this could be a source of variation in the date, since clearly half of the sample lived for less than a year in Germany at the time of the experiment, and half of the sample from 1 to 7 years.

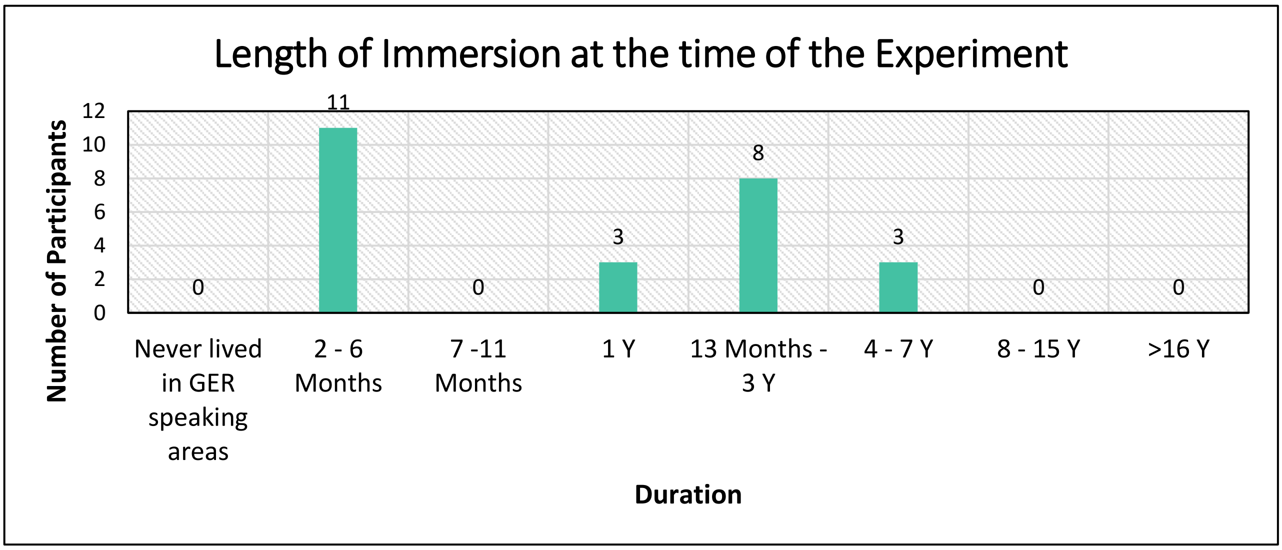


Figure 4.9.4 Length of Immersion at the time of the Experiment

The last questions of the questionnaire pertained the amount of time in which participants averagely were exposed to the auditive perception of German in the form of Audio-visual material, like Movies and general TV, and amount of time they averagely listened to Radio. Moreover, the perceived difficulty of these tasks was asked, in order to have a measure of how hard the experimental manipulation involving only auditive perception without visual information could have been perceived. The distribution of the total amount of hours per week in which participants usually watched audio-visual material like TV and movies is plotted Figure 4.9.5 below.

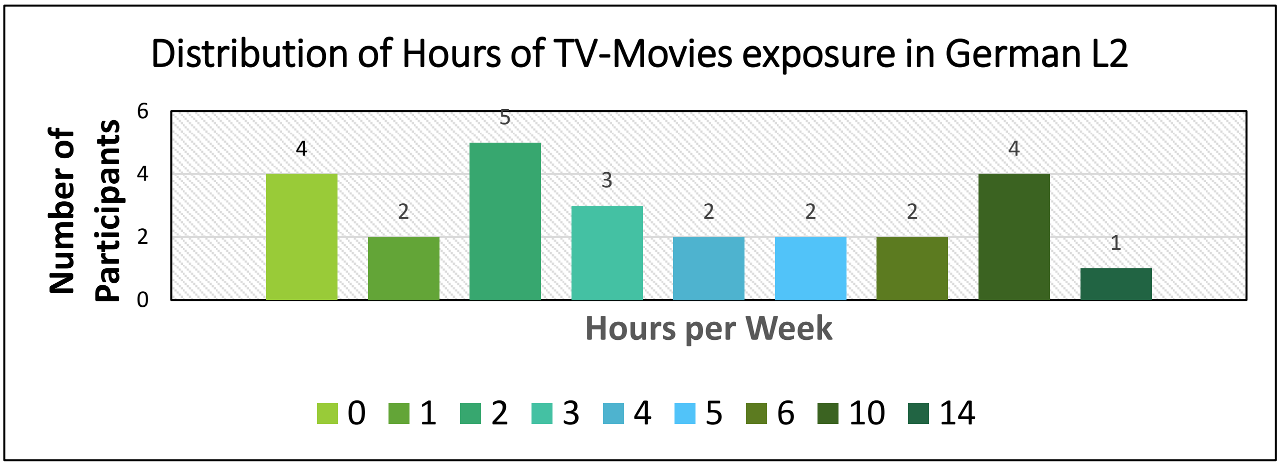


Figure 4.9.5 Distribution of Tv-Movies exposure in German

When asked about the perceived difficulty of watching and comprehending audio-visual material in German L2, of 5 level Likert scale, out the 21 participants who made use of it: 0 replied it is extremely hard, non-recreative, and extremely tiring with a very low level of comprehension; 4 replied they found it slightly recreational but they were not able to continuously concentrate for long periods of time, and that overall they comprehended the general meaning of the presented scenes and dialogues; 14 replied that found it overall a recreational activity, that could be prolonged over long times without being too tiring, and that they comprehended most of the dialogues and scenes, they still found hard to understand jokes, sarcasm, culture specific dialogues; 3 responded that they were able to comprehend full dialogues with ease, that it was an enjoyable activity, that could be prolonged over time, and that they were able to comprehend jokes and sarcasm; 0 participants responded that he/she had the ability to enjoy to watch TV and movies with ease, understanding sarcasm and in addition to be able to identify regional accents in spoken German L2 correctly.

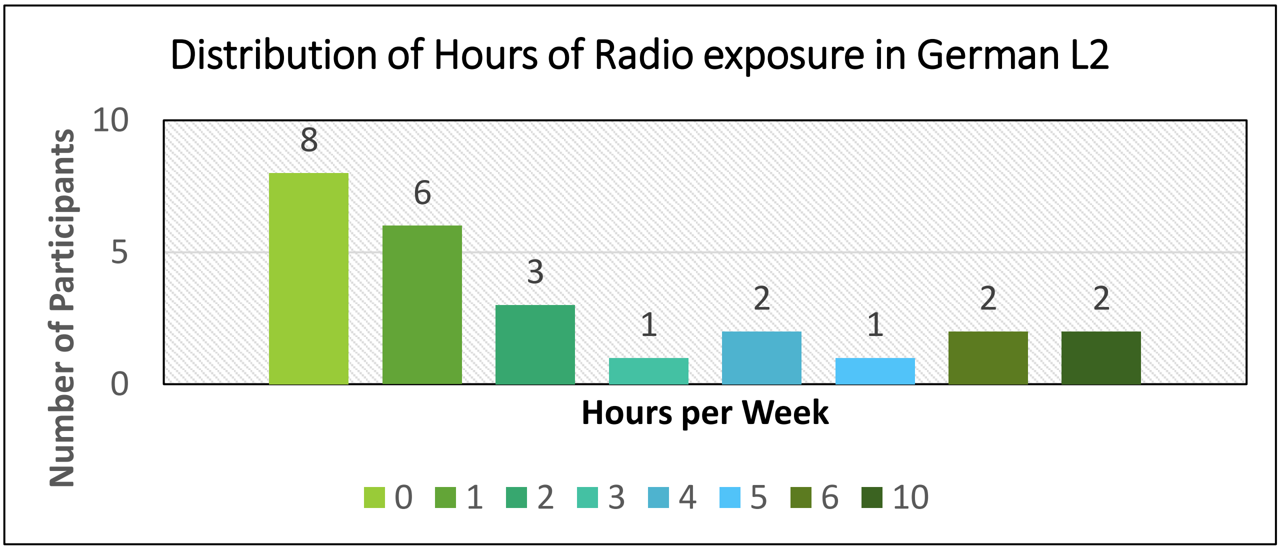


Figure 4.9.6 Distribution of Hours of Radio per week in German L2

The variable measuring the hours that participants spend listening to Radio also divides the sample into two possible clusters, since the majority of the sample showed to listen to Radio for less than 3 hours a week, and on the contrary 5 participants declared from 5 to 10 hours per week, corresponding at al least a minimum of one hour a day. Out of 25 participants, only 2 found the activity of listening to Radio and comprehending speech extremely easy, 13 found it rather easy, and 10 declared they found it rather difficult. This, also, could be a relevant variable where to investigate possible covariates in the experiment.

4.10. Referring ERP response in the original study of Schumacher and Baumann (2010)

Single-Electrodes plot adapted from Baumann & Schumacher (2010)

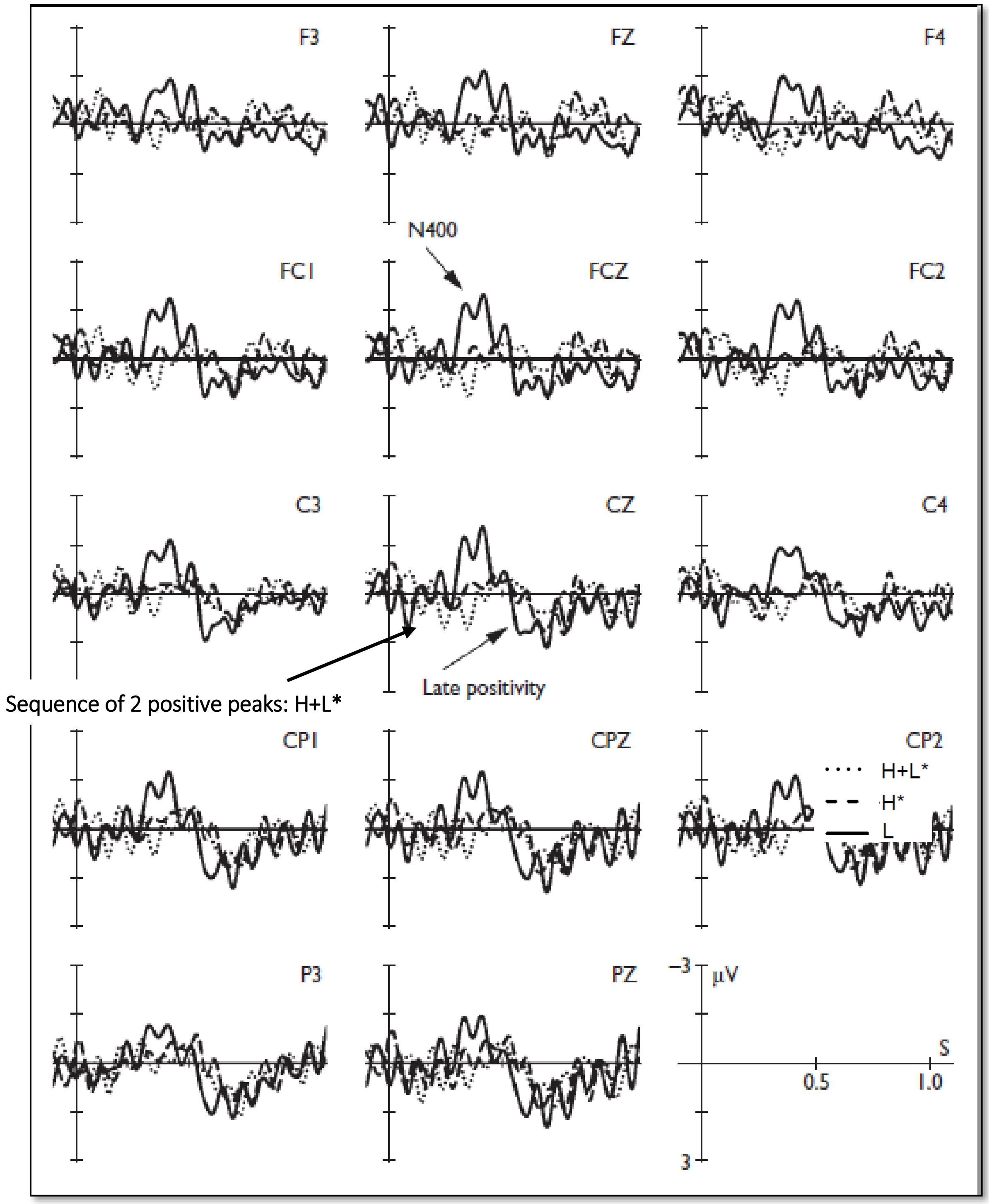


Figure 4.10.1 Biphasic pattern (adapted from Baumann & Schumacher, 2010)

4.11. Results

Analyses are carried out with the trigger timed-locked to the onset of the Target NP in the second sentence. Firstly, I report the analysis of L1 speakers done with the same procedure as in Schumacher & Baumann (2010), in order to observe whether the results replicate the previous ones with a similar manipulation or any significant difference appears. The time windows investigated are the ones adopted in the previous paper, thus: 320-440 ms (N400 window) and 520-640 ms (Late Positivity window); the topographic distribution also is analogous to the one adopted in the previous study and is composed of a midline region (ch: Fz, FCz, Cz, CPz, Pz, POz) and a lateral factor having 4 levels: left anterior (F3/F7/FC1/FC5), right anterior (F4/F8/FC2/ FC6), left posterior (P3/P7/CP1/CP5), right posterior (P4/P8/CP2/CP6. Analyses are carried out hierarchically on the mean amplitude value per condition in each time window, followed by pairwise comparisons.

Single-Electrodes plot in the current study – 20 participants native speakers of German

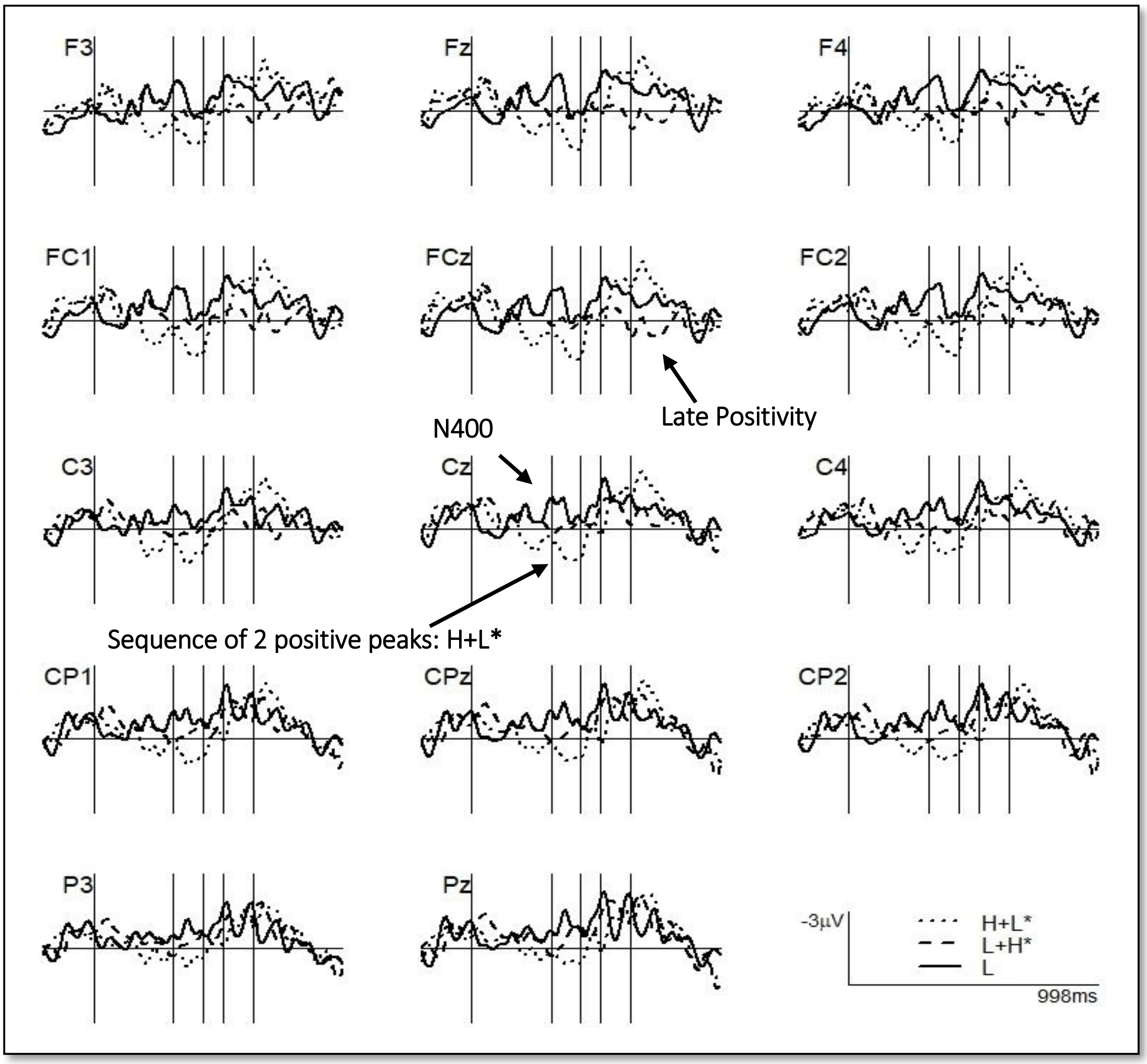


Figure 4.11.1 L1 German speakers

Single-Electrodes plot in the current study – 25 participants Italian L1, German L2 speakers

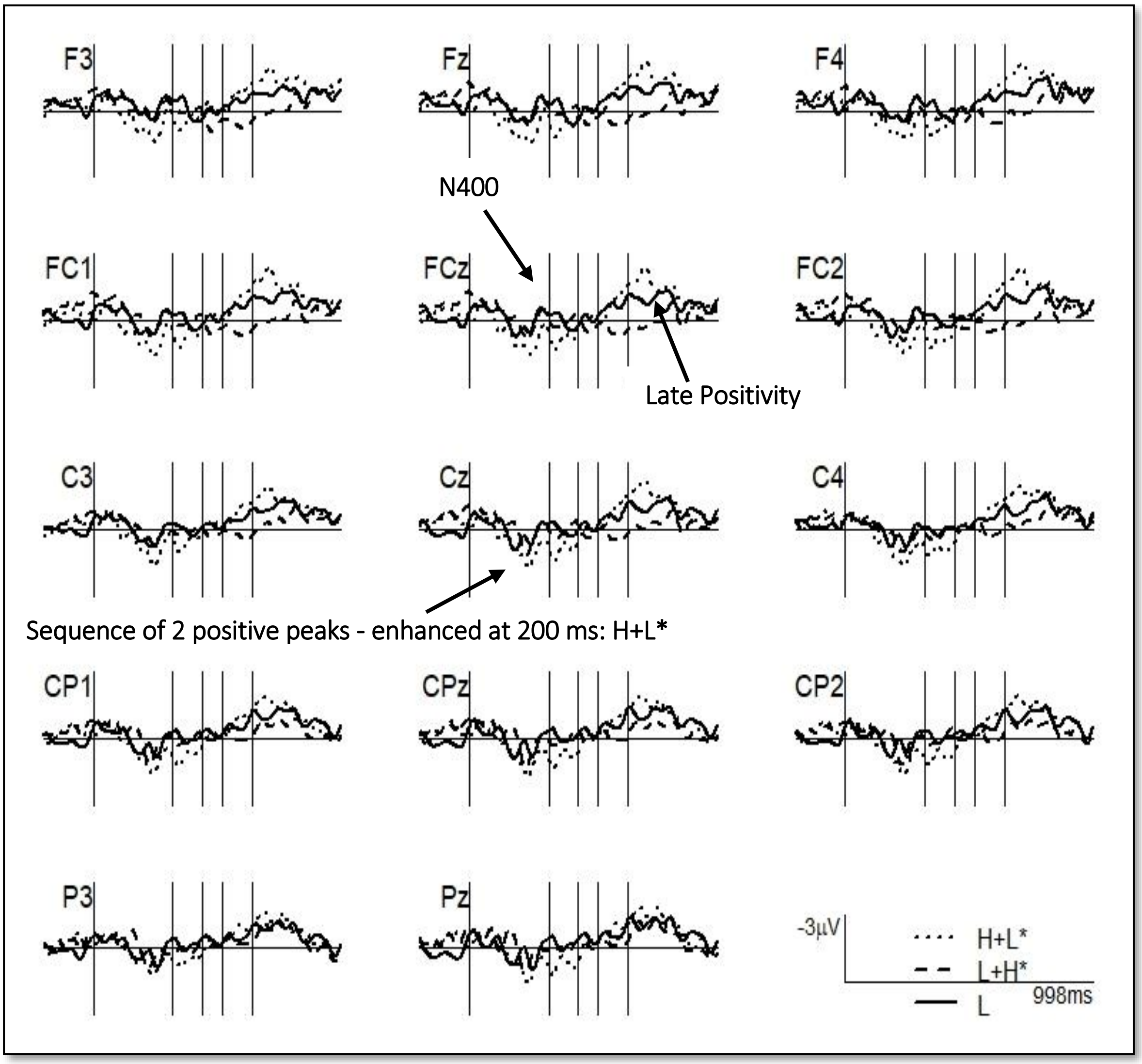


Figure 4.11.2 L2 German speaker (L1 Italian)

4.11.1. Replication of Schumacher & Baumann (2010)

In L1 speakers:

Schumacher & Baumann (2010) found in both time windows (N400: from 320 – 440 ms; Late-Positivity: 520 – 640) a main effect of the 3 levels factor PROSODY over the midline electrodes.

In the 320 – 440 ms time window, the current data replicate the effect for PROSODY over the midline (F (2,38) = 13.85; p < 0.001) and, moreover, the current data show a significant topographic effect over the 6 electrodes F (5, 95) = 3.80; p < 0.01), but no significant interaction.

On the other hand, in the later time window, the current data do not replicate a main effect for PROSODY over the midline, but they again show a topographic effect of the electrodes (F (5, 95) = 5.19; p < 0.001) and they replicate the significant interaction (F (10, 190) = 12.11; p <0.00001).

Over the remaining electrodes, organized in a lateral topographic factor, Schumacher & Baumann (2010) found, respectively, in the 320-440 ms time window a main effect of PROSODY, and in the 520-640 ms a significant interaction between PROSODY and the topographic factor.

Over lateral electrodes, in the 320-440 ms time window, the current data replicate the a main effect of PROSODY (F(2, 38) = 14.26; p < 0.0001); and they show a significant topographic effect (F(3, 57) = 4.32; p < 0.01); in the 520-640 ms time window, the data show a non-significant tendency for PROSODY (F (2,38) = 2.84; p = 0.07); a significant topographic effect (F (3,57) = 5.17; p < 0.01) and they replicate the interaction between PROSODY and the topographic factor (F (6,114) = 10.79; p< 0.00000001).

In Schumacher & Baumann (2010), pairwise comparisons over the midline in the 320-440 ms time window, showed a significant effect of PROSODY across all 3 contrasts; and in the 520 – 640 ms time window, a significant mean difference in the comparison between the deaccenting condition and H*, and the deaccenting condition vs. the congruent H+L* condition; they found no difference between the congruent and H* accent.

In the current study, pairwise comparisons over the midline in the 320-440 ms time window are all significant, and FDR corrected for 3 comparisons. These results fully replicate the previous study. In the later time window, the current data do not show any significant comparison; therefore, the current data do not show a particular positivity for the deaccenting condition.

L1 (Replication Analysis) Midline (320 – 440 ms)							
Contrast		t	df	Mean difference	p.value	FDR adjusted p.value	Confidence Interval
1	H+L* - Deacc	5.0022	19	1.5706	7.9110 e-05	0.0002	0.9135 2.2278
2	H+L* - Contrastive	3.2797	19	0.8995	3.9414 e-03	0.0059	0.3255 1.4735
3	Contrastive - Deacc	-2.1745	19	-0.6712	4.2503 e-02	0.0425	-1.3171 -0.0252

Table 4.11.1 Replication: Post - Hoc T-tests N400 window over the Midline in L1 German speakers

L1 (Replication Analysis) Midline (520 – 640 ms)							
Contrast		T	df	Mean difference	p.value	FDR adjusted p.value	Confidence Interval
1	H+L* - Deacc	1.0005	19	0.3091	0.3296	0.4945	-0.3376 0.9559
2	H+L* - Contrastive	-0.3128	19	-0.0948	0.7578	0.7578	-0.7291 0.5395
3	Contrastive - Deacc	-1.4630	19	-0.4040	0.1598	0.4794	-0.9819 0.1740

Table 4.11.2 Replication: Post - Hoc T-tests Late Pos window over the Midline in L1 German speakers

In Schumacher & Baumann (2010), pairwise comparisons over the lateral topographic factor, in the 320-440 ms time window, highlighted a significant mean difference between the deaccenting condition and the two other conditions, and a non-significant difference between the newness accent and the congruent H+L* accent. In this study all 3 comparisons, FDR corrected for 3 contrasts, are significantly different.

L1 (Replication Analysis) Lateral Factor (320 – 440 ms)							
Contrast		t	df	Mean difference	p.value	FDR adjusted p.value	Confidence Interval
1	H+L* - Deacc	4.8165	19	1.1316	0.0001	0.0004	0.6399 1.6234
2	H+L* - Contrastive	3.2599	19	0.5936	0.0041	0.0062	0.2125 0.9747
3	Contrastive - Deacc	-2.4963	19	-0.5381	0.0219	0.0219	-0.9892 -0.0869

Table 4.11.3 Replication: Post - Hoc T-tests N400 window over the Lateral electrodes in L1 German speakers

In the late positivity time window, Schumacher & Baumann (2010) found, over lateral electrodes, analogous results as over the midline (no significant difference between accessibility and newness access).
The current results as over the midline are analogous to the ones from the originally study, but no significant comparison is found.

L1 (Replication Analysis) Lateral Factor (520 – 640 ms)							
Contrast		t	df	Mean difference	p.value	FDR adjusted p.value	Confidence Interval
1	H+L* - Deacc	0.9271	19	0.1941	0.3655	0.3655	-0.2441 0.6322
2	H+L* - Contrastive	-1.5145	19	-0.3152	0.1464	0.2195	-0.7509 0.1204
3	Contrastive - Deacc	-2.2257	19	-0.5093	0.0383	0.1150	-0.9882 -0.0304

Table 4.11.4 Replication: Post - Hoc T-tests Late Pos window over the Lateral electrodes in L1 German speakers

Overall, the results highlight a full replication of the PROSODY effect in the 320-440 ms time window over the midline and both lateral electrodes, as well as in single comparisons. On the contrary, in the late time window, the interaction between PROSODY and the topographic factor over the midline is replicated, but not the main effect. Analogously, the current data do not replicate the significant difference in the pairwise comparisons. In particular, further analysis of the interaction between the PROSODY factor and the topographic factor in the late-positivity window, highlight that pairwise comparisons display a sensibility for the PROSODY factor mainly in the Left Anterior quarter. On the contrary, the main effect found in Schumacher & Baumann (2010) in the late-positivity window emerged in posterior electrodes.
The further difference in late-effects pertains the change and introduction of the different PA (L+H*), possibly signaling contrastive information, that in the left anterior quarter is significantly different from both the congruent PA and the Deaccenting condition, in opposition to the fact that the appropriate accent and the deaccenting condition do not show a significant mean difference.

L1 (Replication Analysis) Anterior Left – L1 (520 – 640 ms)							
Contrast		t	df	Mean difference	p.value	FDR adjusted p.value	Confidence Interval
1	H+L* - Deacc	0.0987	19	0.0279	0.9224	0.9224	-0.5637 0.6195
2	H+L* - Contrastive	-3.5258	19	-0.9934	0.0023	0.0068	-1.5831 -0.4037
3	Contrastive - Deacc	-3.2119	19	-1.0213	0.0046	0.0069	-1.6869 -0.3558

Table 4.11.5 Replication: Post - Hoc T-tests Late Pos window in Anterior Left quarter in L1 German speakers

Processing differences from the original Schumacher & Baumann's experiment also appear in the latencies where the effects are maximal and in the duration of effects. The maximal mean difference among conditions appears to begin slightly earlier from the previously set time-windows and to last longer especially in the later time window; on the contrary, the data the effect is maximal after 640 ms.
Moreover, by visual inspection, it is possible to observe that the congruent accent for accessibility (H+L*) also elicited in the original experiment a double positive peak, one at about 250 ms and the second falling in the N400 time window. In the previous experiment the presence of this early positive peak has not been discussed, and in this experiment, it shows to be more enhanced and present both in native and second language learners' waveforms.
On one side, these observations indicate that the changes in the current manipulation (slower speech-rate and emphasized modulation of F0 contours) compared to the original experiment can have enhanced the predictability of the prosodic manipulation in the 3 conditions, accelerating the recognition of pitch contours and their integration with the matching/non-matching degree of activation of the target referent. On the other one, the slower tempo can have caused the effect that at the latency of 600 ms speech is still being acoustically processed by participants, compared to the original native-like speech rate in Schumacher & Baumann (2010). The still on-going acoustical speech processing can have slowed the processing of discourse updating in the later time window, delaying the effects.

In final and more important instance, the first analysis highlights in the late time window that the change in PA between the two experiments causes the most significant processing difference. It is in fact, the accent conveying contrastiveness (L+H*, dashed/blue line) that correlates with the highest late- positivity, and not the Deaccenting condition.

For these reasons, both L1 and L2 new data will be analyzed adopting three new time-windows that best capture L1 and L2 processing differences independently from the original study.
The first time-window seeks to observe the nature of the first positive peak for the congruent accessibility accent in both language groups. The first time-window is set between 180 – 300 ms.
The second time window is set to highlight the comparison of L1 and L2 speakers' data of the integration of prosody and the degree of activation of the referent, emerging in the N400 component.
The second time window is set between 300 – 440 ms.
The third time window is set to capture the longer duration of the late-positivity effect that in both language groups lasts longer than 640 ms in both language groups, and it is functionally associated to the cost of discourse update.
The third time window is set between 570 – 800 ms.

4.11.2. New Analysis with 3 time-windows: L1 and L2 processing

4.11.2.1. <u>L1</u>

4.11.2.1.1. L1 - Lateral

Over lateral electrodes, in the 180 – 300 ms time window, a main effect of PROSODY $F(2,38) = 10.79$; $p = 0.0002$ is found.
In the 300 – 440 ms time window, there is an effect for Prosody $F(2,38) = 14.57$; $p = 0.00002$; in addition, a topographic effect of the Lateral factor $F(3,57) = 3.91$; $p = 0.02$.
Finally, in the 570 – 800 ms time window, the data show both an effect for Prosody $F(2,38) = 5.87$; $p = 0.0060$; a topographic effect of the Lateral factor $F(3, 57) = 4.65$; $p = 0.0056$; and an interaction with the main prosody effect $F(6,114) = 14.93$; $p = 1.53$ e-12.
Over the lateral factor in L1, pairwise comparisons FDR corrected for 3 within-subjects contrasts, in the 180-300 ms time window, show a significant effect of PROSODY in two comparisons: Congruent vs. Deaccentuation $t(19) = 6.74$; $p = 5.82$ e-06; C.I.(0.58 - 1.10); Congruent vs. Contrastive $t(19) = 3.09$; $p = 0.009$; C.I.(0.21 - 1.11); and no significant mean difference between the Contrastive and Deaccentuation condition.
In the 300 – 440 ms time window, the data show a significant 3-way effect of PROSODY: Congruent vs. Deaccentuation $t(19) = 5.03$; $p = 0.0002$; C.I.(0.63 - 1.54); Congruent vs. Contrastive $t(19) = 2.89$; $p = 0.01$; C.I.(0.14 - 0.89); Contrastive vs. Deaccentuation $t(19) = -2.76$; $p = 0.01$; C.I.(-1.00 - -0.14);
In the 570 – 800 ms time window the factor PROSODY shows a significant mean difference only between the Congruent and Contrastive accent $t(19) = -3.50$; $p = 0.007$; C.I.(-1.19 - -0.30); in addition, there is a non-significant tendency in the Congruent vs. Deaccentuation condition, $t(19) = -2.06$; $p = 0.08$; C.I.(-0.81 - 0.01); and no significant mean difference in Contrastive vs. Deaccentuation.

4.11.2.1.2. L1 - Midline

Over the midline, in the 180 – 300 ms time window, the data show an effect for PROSODY F (2,38) = 5.89; p = 0.0059; and an interaction between the topographic effect over the 6 electrodes and the main prosody effect F(10, 190) = 3.64; p = 0.0002.

In the 300 – 440 ms time window, an effect for Prosody is present F(2,38) = 13.92; p = 0.000029; in addition, there is also a topographic effect over the 6 electrodes F(5, 95) = 3.14; p = 0.014.

In the 570 – 800 ms time window, the effect of Prosody is not significant, but topographic effect over the 6 electrodes is present F(5,95) = 9.03; p = 4.76 e-07; and an interaction between the topographic factor with the main prosody effect F(10, 190) = 18.87; p = 8.46 e-24.

Over the midline in L1, pairwise comparisons FDR corrected for 3 within-subjects contrasts, in the 180-300 ms time window, show a significant effect of PROSODY in two comparisons: Congruent vs. Deaccentuation t (19) = 4.36 ; p = 0.001; C.I.(0.46 - 1.32); Congruent vs. Contrastive t (19) = 2.45 ; p = 0.04; C.I.(0.12 - 1.56); and no significant mean difference between the Contrastive and Deaccentuation condition.

In the 300 – 440 ms time window, there is a significant 3-way effect of PROSODY: Congruent vs. Deaccentuation t (19) = 5.16 ; p = 0.0002 ; C.I.(0.90 - 2.13); Congruent vs. Contrastive t (19) = 2.99 ; p = 0.02; C.I.(0.24 - 1.38); Contrastive vs. Deaccentuation t (19) = -2.37 ; p = 0.03; C.I.(-1.32 - -0.08);

In the 570 – 800 ms time window the factor PROSODY is not significant in any comparison; only in the Congruent vs. Contrastive comparison a tendency is t(19) = -2.35; p = 0.09; C.I. (-1.46 - -0.08).

In the 570-800 ms time-window, given the significant effect of both Prosody, topographic factor, and the interaction between the two, over lateral electrodes, and the significant effect only of the topographic factor and the interaction over the midline, single comparisons looking for topographic differences in frontal vs. posterior electrodes are investigated.

In the frontal region, the two frontal quarters (Anterior Left + Anterior Right) are collapsed with the two frontal electrodes in the midline (Fz, FCz) and in the posterior region the two posterior quarters (Posterior Left + Posterior Right) are collapsed with the two posterior electrodes in the midline (CPz, Pz).

L1 speakers' results highlight, in the time window of 570-800 ms, that the factor PROSODY displays a 3-way significant effect in the frontal region reported in Table X, and that on the contrary no-significant effect is present in posterior electrodes, as shown in Table 4.11.6 and Table 4.11.7 below.

L1 Frontal (570 – 800 ms) – (AL + AR + Fz + FCz)							
Contrast		t	df	Mean difference	p.value	FDR adjusted p.value	Confidence Interval
1	H+L* - Deacc	-2.2327	19	-0.4863	0.0378	0.0378	-0.9423 -0.0304
2	H+L* - Contrastive	-4.5251	19	-1.3859	0.0002	0.0007	-2.0269 -0.7449
3	Contrastive - Deacc	-2.6800	19	-0.8996	0.0148	0.0222	-2.0269 -0.7449

Table 4.11.6 New Analysis - New Late Pos window over frontal electrodes in L1 German speakers

L1 Posterior (570 – 800 ms) – (PL + PR + CPz + Pz)							
Contrast		t	Df	Mean difference	p.value	FDR adjusted p.value	Confidence Interval
1	H+L* - Deacc	-1.6938	19	-0.3787	0.1066	0.3200	-0.8467 0.0893
2	H+L* - Contrastive	-1.0062	19	-0.2049	0.3270	0.4462	-0.6312 0.2214
3	Contrastive - Deacc	0.7779	19	0.1738	0.4462	0.4462	-0.2938 0.6414

Table 4.11.7 New Analysis – Post-Hoc T-tests - Late Pos window over posterior electrodes in L1 German speakers

4.11.2.2. L2

4.11.2.2.1. L2 - Lateral

Over lateral electrodes, in the 180 – 300 ms time window, the data show a main effect for PROSODY F(2,48) = 10.12; p = 0.0002.

In the 320 – 440 ms time window, also an effect for Prosody is present F(2,48) = 6.40; p = 0.0034.

In the 570 – 800 ms time window, it present both an effect for Prosody F(2,48) = 25.70; p = 2.59 e-08; a topographic effect over lateral electrodes F(3, 72) = 3.72; p = 0.0152 ; and an interaction with the main prosody effect F(6, 144) = 11.52; p = 1.65 e-10.

Over the lateral factor in L2, pairwise comparisons FDR corrected for 3 within-subjects contrasts, in the 180-300 ms time window, show a significant effect of PROSODY in two comparisons: Congruent vs. Deaccentuation t (24) = 3.92; p = 0.002; C.I.(0.29 - 0.93); Congruent vs. Contrastive t (24) = 3.67; p = 0.002; C.I.(0.25 - 0.88); and no significant mean difference between the Contrastive and Deaccentuation condition.

In the 300 – 440 ms time window, there is a significant effect of PROSODY in 2 comparisons: Congruent vs. Deaccentuation t (24) = 3.78; p = 0.003; C.I.(0.27 - 0.91); Congruent vs. Contrastive t(24) = 2.39; p = 0.04; C.I.(0.06 - 0.80); and no significant mean difference in Contrastive vs. Deaccentuation.

In the 570 – 800 ms time window the factor PROSODY shows a 3-way significant effect: Congruent vs. Deaccentuation t (24) = -2.89; p = 0.008; C.I.(-0.72 - -0.12); Congruent vs. Contrastive t (24) = -7.16; p = 6.89 e-07; C.I.(-1.30 -0.72); Contrastive vs. Deaccentuation t(24) = -4.31; p = 0.0004; C.I.(-0.87 - -0.31).

4.11.2.2.2. L2 - Midline

Over the midline, In the 180 – 300 ms time window, a main effect of PROSODY F(2,48) = 8.03; p = 0.001 is found;

In the 320 – 440 ms time window, the main effect of Prosody F(2,48) = 6.76; p = 0.0026 is confirmed.

In the 570 – 800 ms time window, both an effect for Prosody F(2, 48) = 18.62; p = 0.000001; a topographic effect over the 6 electrodes F(5,120) = 3.04; p = 0.0013, and an interaction with the main prosody effect F(10, 240) = 13.85; p = 3.22 e-19 are found.

Over the midline in L2, pairwise comparisons FDR corrected for 3 within-subjects contrasts, in the 180-300 ms time window, show a significant effect of PROSODY in two comparisons: Congruent vs. Deaccentuation t (24) = 3.59; p = 0.004; C.I.(0.3093 1.1456); Congruent vs. Contrastive t (24) = 3.31; p = 0.004; C.I.(0.2822 1.2189); and no significant mean difference between the Contrastive and Deaccentuation condition.

In the 300 – 440 ms time window, there is a significant 2-way effect of PROSODY: Congruent vs. Deaccentuation t (24) = ; p = 0.003 ; C.I.(0.3909 1.3209); Congruent vs. Contrastive t (24) = 2.55; p = 0.03; C.I.(0.1284 1.2258); no significant mean difference in Contrastive vs. Deaccentuation.

In the 570 – 800 ms time window the factor PROSODY shows a significant 3-way effect: Congruent vs. Deaccentuation t (24) = -2.30; p = 0.03; C.I.(-0.89 - -0.05); Congruent vs. Contrastive t (24) = -6.02; p = 9.88 e-06; C.I.(-1.59 -0.78); Contrastive vs. Deaccentuation t(24) = -.86; p = 0.001; C.I.(-1.09 - -0.33).

In L2, in the 570-800 ms time-window, an interaction with the topographic effect is also present, but the single comparisons show to be overall 3-way significant. By visual inspection, nonetheless, the effect appears to be mainly frontal also in L2, therefore the analogous frontal and posterior regions are collapsed and compared in order to verify whether the significant 3-way contrast is caused only by the large effect in the frontal electrodes.

L2 speakers' results highlight , in the time window of 570-800 ms, that the factor PROSODY also displays a strong a 3-way significant effect in the frontal region, but that differently from L1 speakers, two comparisons (congruent vs. Deaccentuation; and Contrastive vs. Deaccentuation), even if the effect is small, are significant, and the comparison of Congruent vs. Contrastive shows to be largely significant also in posterior electrodes.

Results are reported in Tables 4.11.8 & Table 4.11.9.

L2 Frontal (570 – 800 ms) – (AL + AR + Fz + FCz)							
Contrast		t	df	Mean difference	p.value	FDR adjusted p.value	Confidence Interval
1	H+L* - Deacc	-2.817	24	-0.5395	0.0096	0.0096	-0.9348 -0.1442
2	H+L* - Contrastive	- 7.5462	24	-1.5212	8.73 e-08	2.62 e-07	-1.9372 -1.1051
3	Contrastive - Deacc	-4.7567	24	-0.9817	7.71 e-05	0.0001	-1.4076 -0.5557

Table 4.11.8 New Analysis - New Late Pos window over frontal electrodes in L2 German speakers

L2 Posterior (570 – 800 ms) – (PL + PR + CPz + Pz)							
Contrast		t	Df	Mean difference	p.value	FDR adjusted p.value	Confidence Interval
1	H+L* - Deacc	-2.1565	24	-0.3284	0.0413	0.0434	-0.6426 -0.014
2	H+L* - Contrastive	-4.3885	24	-0.6138	0.0002	0.0006	-0.9025 -0.3252
3	Contrastive - Deacc	-2.1326	24	-0.2855	0.0434	0.0434	-0.5618 -0.0092

Table 4.11.9 New Analysis - New Late Pos window over posterior electrodes in L2 German speakers

Single-Electrodes plot with New Time-windows – 20 participants native speakers of German

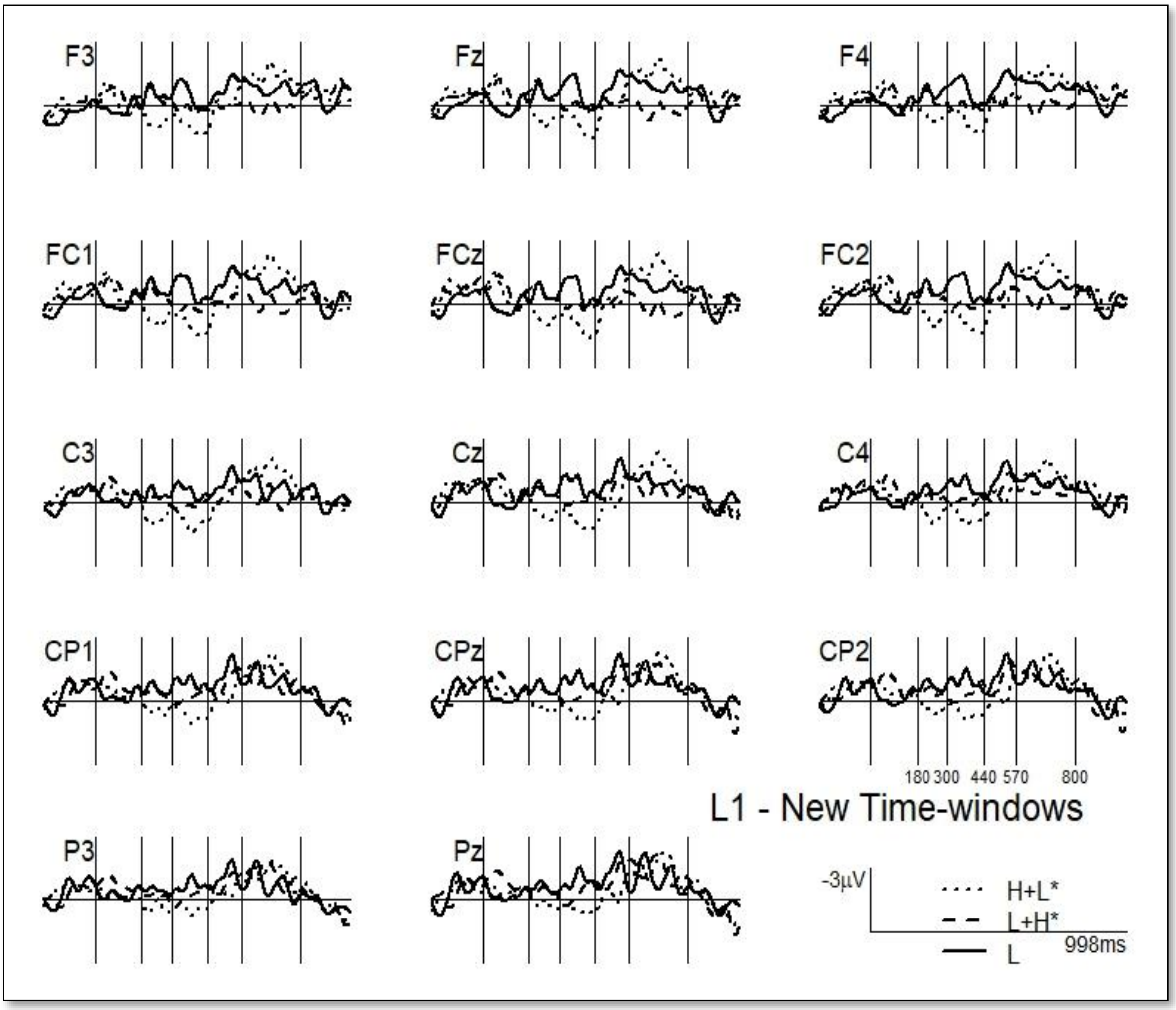


Figure 4.11.3 L1 results with new time-windows

Single-Electrodes plot with New Time-Windows – 20 participants native speakers of German

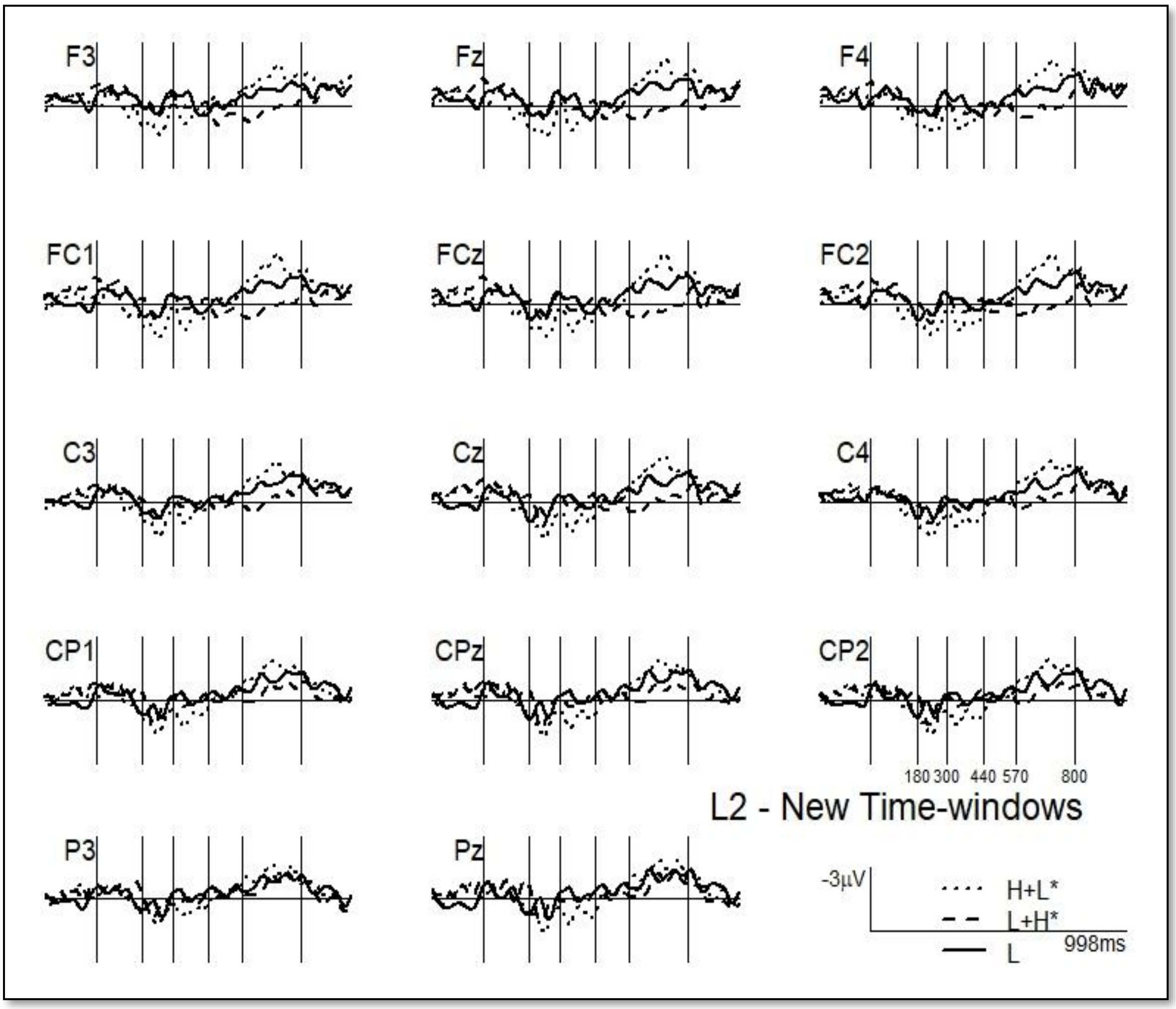


Figure 4.11.4 L2 results with new time-windows

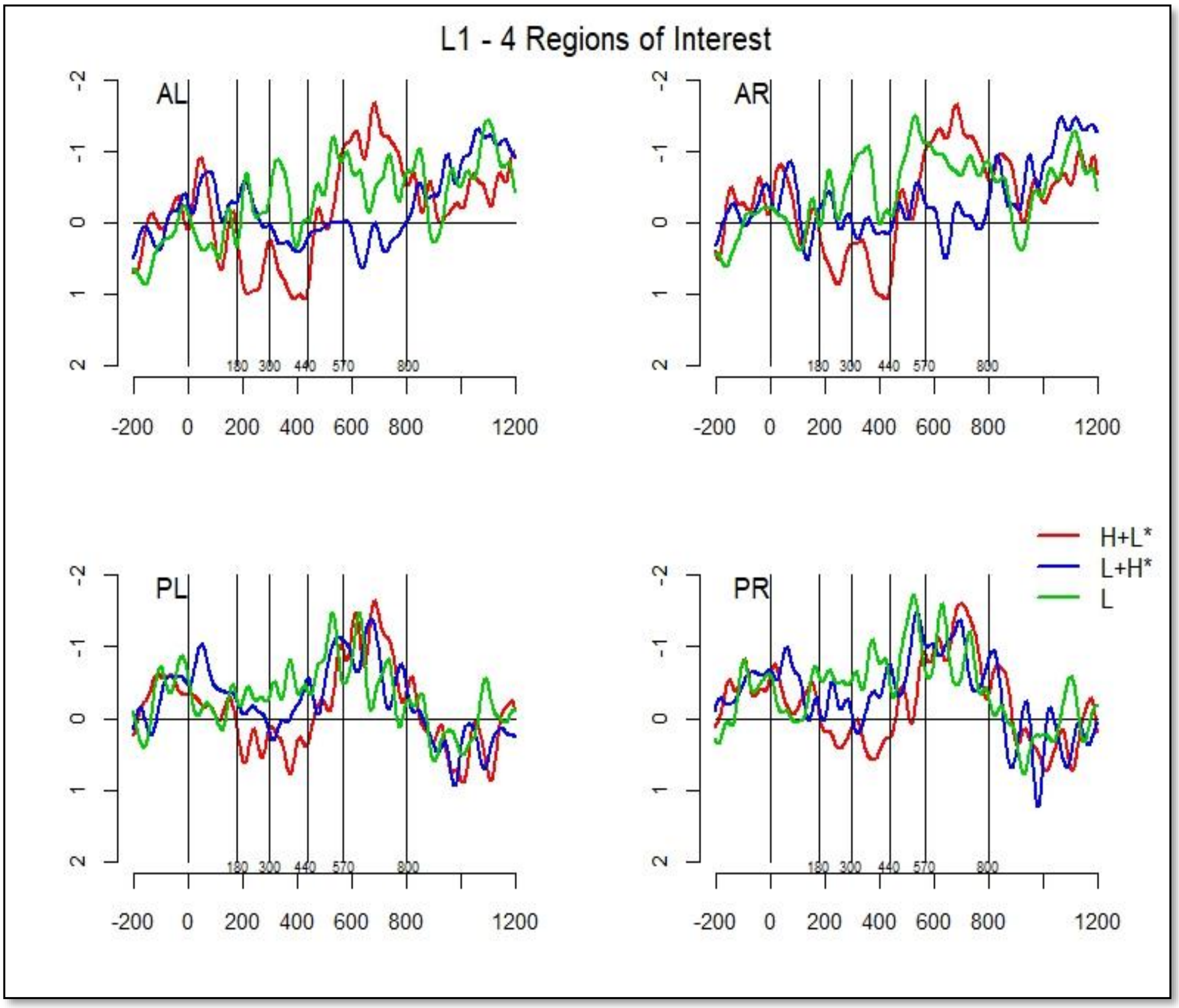


Figure 4.11.5 L1 results aggregated in 4 ROIs in new time-windows

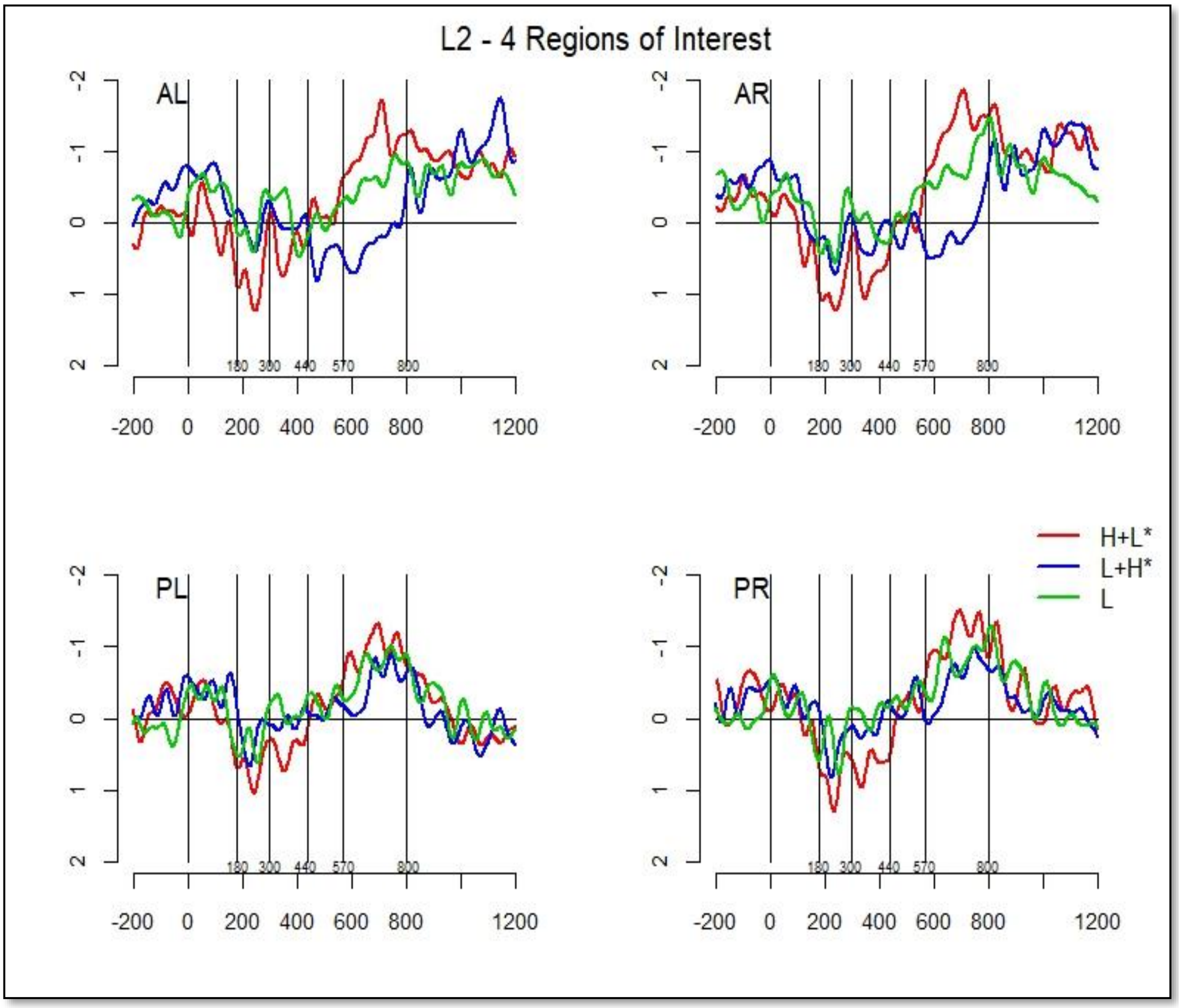


Figure 4.11.6 L2 results aggregated in 4 ROIs in new time-windows

4.11.2.3. L1 vs. L2

In the 180 – 300 ms time window, comparisons of single accent's un-subtracted waves in each topographic quarter, elicited by L1 and L2 speakers, highlight: no significant difference in the Anterior Left quarter in any accent's comparison, a significant difference in the Deaccentuation condition both in the Anterior Right quarter $F(1,43) = 4.20$; $p = 0.047$; and in the Posterior Left quarter $F(1,43) = 4.23$; $p = 0.006$.
In the Posterior Right quarter all 3 accents' waveforms are significantly different between L1 and L2 speakers: Congruent (H+L*): $F(1,43) = 26.68$ $p = 5.46$ e-06; Contrastive Accent $F(1,43) = 6.46$ $p = 0.015$; and Deaccentuation condition: $F(1,43) = 13.32$ $p = 0.0007$.

In the 300 – 440 ms time window, no significant difference in the Anterior and Posterior Left quarter is found in any accent's comparison; in the Anterior and Posterior Right quarter a significant difference between L1 and L2 waveforms for the Deaccentuation condition is found, respectively AR: $F(1,43) = 7.66$; $p = 0.008$; PR: $F(1,43) = 9.22$; $p = 0.004$.

In the 570-800 ms time window there is no significant difference in any quarter for any accent-comparison.

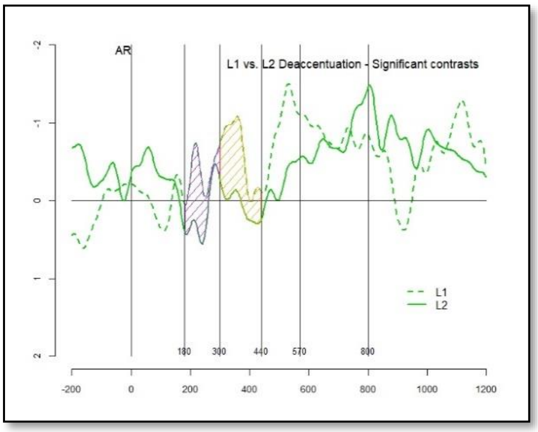


Figure 4.11.7 L1 vs. L2 Deaccentuation's un-subtracted waveforms in AR - ROI

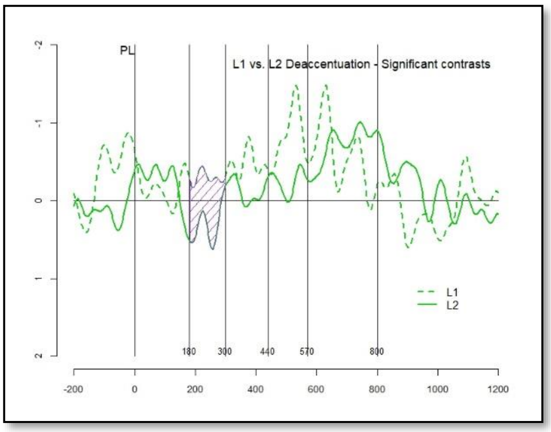


Figure 4.11.8 L2 vs. L2 Deaccentuation's un-subtracted waveforms in PL - ROI

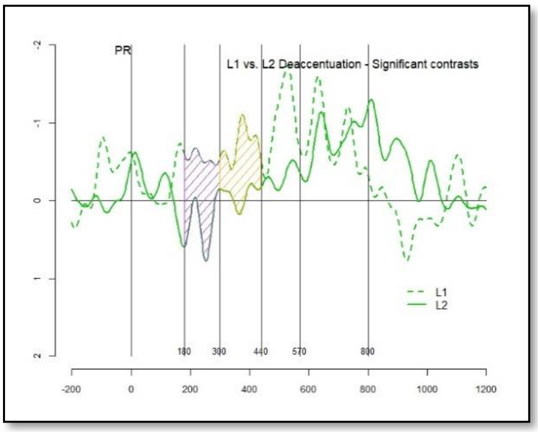


Figure 4.11.9 L1 vs. L2 Deaccentuation's un-subtracted waveforms in PR - ROI

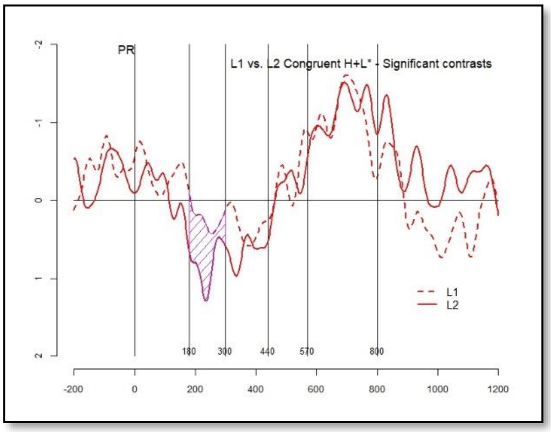


Figure 4.11.11 L1 vs. L2 Congruent-condition's un-subtracted waveforms in PR - ROI

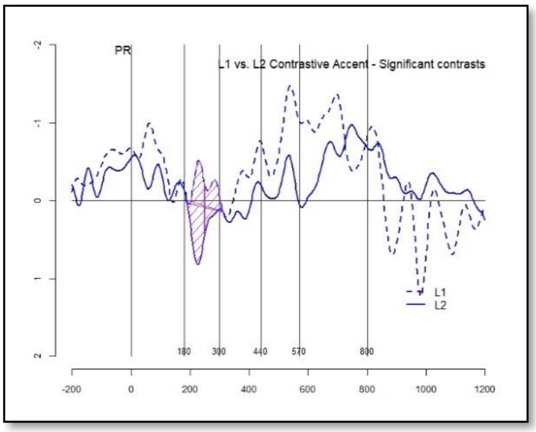


Figure 4.11.10 Figure 4.10.11 L1 vs. L2 Contrastive condition's un-subtracted waveforms in PR - ROI

4.11.3. Discussion of group effects

4.11.3.1. Replication of Schumacher and Baumann

The ERP analysis performed adopting the same parameters and specifics of Schumacher & Baumann (2010) highlighted that the 3 accents aligned in this experiment (L+H* - incongruent, blue-dashed line; H+L* - congruent, red-solid line; and Deaccentuation pattern – incongruent, green-dotted line) fully replicate the main PROSODY effect in the N400 window (330-440 ms) and interact with the topographic factor showing a larger effect in frontal electrodes. A further analysis of the two accents comparison in the same time-window highlights a replication of the 3-way modulation of the N400 component. On the other hand, adopting the late time-window set as in Schumacher & Baumann (2010), thus 520-640, the main effect for PROSODY does not appear, and it appears an interaction with the topographic factor signaling also, in this case, a larger frontal effect. The late-positivity component in the original study was mainly posterior and associated with the Deaccentuation pattern. In this study, the larger positivity effect is anterior and mainly responding to the newly introduced peak L+H*.
By visual inspection, it is also noticeable that the main late-positivity effect begins later and lasts longer than in the original study, and that the congruent condition (red-solid line) elicited two early positive peaks, one between 180-300 ms and one in correspondence of the N400 window, modulating the un-subtracted ERP component with the exact opposite polarity of the two incongruent conditions that are classified as negativities in the N400 window, and also display a smaller negative effect in the 180.300 ms time window. This positive modulation before 440 ms of the congruent condition was present also in the results of the original study but with a smaller effect and therefore it had not been further discussed.

The replication analysis shows that the slowing tempo of the speech produced by the professional speaker elicited a more enhanced early processing, that appears to be consistent with the congruency of the pitch modulation with the degrees of mental activation of the referent and did not affect the modulation of the N400. The 3 pitch contours adopted in this study, therefore, are associable with a 3-way modulation of the integration cost with the previously processed context. Following the interpretation provided by Schumacher & Baumann (2010), the modulation of the negativity, higher for Deaccentuation, followed by the steep F0 peak (L+H*), and then by the congruent condition, reflects the increase of difficulty in the integration of discourse when prosody and information status mismatch.

> "Regarding the N400, prominence relations are computed during discourse linking that reflect the referent's accessibility and support its integration with the prior discourse: the more difficult the integration, the more enhanced is the N400 amplitude" (Schumacher and Baumann, 2010; p.3)

On one side, these results support the view that the prosodic modulation is a crucial factor in the computation of the referent's accessibility in German, and that the N400 component, sensible to the cost of the integration of semantic relations, is also sensible to pragmatic and prosodic manipulations. The negativity cannot be interpreted as an Expectancy Negativity (EN) (Heim & Alter, 2004) since the experimental manipulation did not adopt the question-answer paradigm. The type of PA and either broad or narrow focus cannot be expected on the basis of the context sentence.

On the other side, if the N400 component is discussed considering both the integration-view vs. the access-view, introduced in the introduction of this chapter, it appears that in the current results, the presence of a negativity in an early time-window at about 250 ms, hypothesized by Nieuwland et al., (2019) as the indicator for a specific predictability effect, is not found. Instead, the current results indicate that the most predictable accent, on the basis of its phonological form, and its mapping with previously activated representation, elicited a positivity instead, in the same time-window.

The interpretation of the N400 modulation as the representation of the cost of integration, as previously claimed in the original study, appears more plausible also within the broader perspective adopted in the whole thesis. Namely, that the processing of the Target NP entails information coming from both the sensory decoding of the incoming stimulus and the presence of a pre-activated structure given by the referential accessibility of the discourse. In this sense, the N400 appears to reflect the integration of the converging levels: sound domain, semantic relationships, congruent syntactical structure, and fitting discourse model. The N400, therefore, can be an indicator of what within the Firthian Prosodic Analysis, and Polysp model (Hawkins, 2003; in section 1.3) is termed "perceptual coherence."

Concerning the late positivity, the current manipulation shows significant differences from the positivity found by Schumacher & Baumann (2010), since the PROSODY effect was not significant in the proposed time-window, but begins later, and it is mainly anterior.

It is possible that the different PA, highly acoustically salient and signaling a narrow-focused unit, affected the late processing where the discourse is being updated, and the introduction of the new referent is assumed to be performed. It can be hypothesized that the introduction of an incongruent narrow contrastive focus indeed still causes and a higher effort during the reorganization of the discourse. Firstly, because the extremely high peak is incongruent with the semi-active status of the referent, as H* was. In addition, in this late component and in this specific condition, the focus manipulation can be summed to the manipulation of Givenness within the component. Focused units are reported to elicit a P600 component explained as the increased cognitive resources allocated in response to discourse and prosodic prominence (Bornkessel, 2003). The instruction of contrasting the referent with previously mentioned information requires the evaluation of the processed discourse in memory, the linking with the anchor

within the bridging inference, the "whole" concept, and the evaluation of the incongruence since the information carried by the target referent is not contrasting its anchor. The violation of the Givenness modulation can then be added to the higher cognitive effort of memory retrieving, anchor identification and matching – non-matching evaluation.
This process might require a longer time, and this could explain the later developing of the effect and its longer duration. Within this interpretation, the L+H* prosodic contour conveys a communicative instruction of a different nature from the one artificially and semantic created in the stimuli, namely the modulation of pitch accents in function of Givenness matching or non-matching the intermediate level of the target referent. Instead, it manipulates the communicative intention of the target sentence, narrowing the focus, and therefore causes an enhancement of the attention allocation towards the single final referent and communicates that it somewhat carries incongruous information. This instruction added and aligned to the final referent might require a deeper effort when compared to a pitch contour signaling completely active information, but not a contrast relevant to the hearer. The positivity is much larger for the F0 peak conveying the narrow focus than for Deaccentuation.

In second instance, the slower pitch tempo, as highlighted by acoustical analysis, causes that after 500 ms the listeners are still perceiving speech, since the analysis is time-locked to the beginning of the NP, therefore on the preceding article. This means that in the late-window, it is possible to catch the auditive evaluation of the stressed syllable of the target referent, that in the high peak L+H* condition is much acoustically salient than the two other manipulations and represents the main difference from the first Schumacher & Baumann (2010) experiment.

Since the processing of focused elements has been found to elicit early positivity components, such as the P200, linked to higher attention allocation, it is possible that in the broad response to the general discourse evaluation and reorganization observed in the late-positivity, there is the interaction with the auditive response to the highly acoustically salient peak, catching the attention of the participants. The topographic distribution shows to be in accordance with this explanation since all effects are mainly anterior, compatible with the requirement of attentional mechanisms and the evaluation of pitch movements (Hickock & Poeppel, 2007).

The overall positivity modulation, on the other hand, does not seem to fully resemble the positivity effect found for the expected, focused unit in experiments only manipulating the association between accentuation and focus without precisely addressing the manipulation of the degrees of activation. This type of positivity is also reported in the early response between 200 ms (Dimitrova et al. 2012). The auditive response that preliminary emerges in these data and is discussed below, on the other hand, appears to be sensitive to the Givenness modulation, since the larger positivity is associated to the congruent PA, and not the most acoustically salient focused one. The following N400 component then fully indicates that the Givenness modulation is processed and it is the leading mechanism guiding the native-speakers' processing. The hypothesis that the late-positivity related to the L+H* reflects a Close-Positive-Shift component (CPS) is also discarded since the EEG waveform is time-locked to the article, the CPS is normally found from 500 ms of the target word and therefore should be out from the selected epoch. I support therefore the thesis that the late positivity in this study, in the group of L1 speakers, reflects a higher effort in the discourse updating caused by the sum of 3 different factors: a pitch modulation not matching the contextual processed information in terms of degrees of Givenness, a non-matching instruction to identify a contrasted referent in the previous context increasing the memory load, and a higher attentional allocation associated to the processing of a high acoustically salient F0 peak on the stressed syllable of the target referent.

In the following section, I address the L1 vs. L2 group comparison discussing 3 newly set time windows. I consider therefore these results completely independent from the original Schumacher & Baumann (2010) study.

4.11.3.2. Discussion of L1 vs. L2

On the basis of the first replication analysis I have set the following relevant time-windows: early component: 180-300 ms; N400 window: 300-440 ms; Late-positivity window: 570-800 ms. The time windows have been adopted both for L1 and L2 groups and for their comparisons.

4.10.3.2.1. L1

In the L1 group, the analysis confirmed what previously discussed also on the basis of visual inspection. A main effect of prosody in the 2 early time windows and an interaction with the topographic factor in the N400 window. The PROSODY effect in the 180-300 ms time window is broadly distributed, and single comparisons highlight a two-way modulation, of congruent vs. incongruent accentuation pattern. In the N400 window, the 3-way modulation is confirmed. In the late time window, the largest effect is granted by the L+H* high F0 peak modulation and the main effect is two-way, contrastive accent vs. congruent and Deaccentuation. When observing a partition of the scalp, given the interaction with the topographic factor, the effect is 3-way in the anterior electrodes, and absent in the posterior ones.

This supports the previously discussed results, confirming a strong interplay, non-investigated in the previous experiment, between required attention allocation during the auditive perception and the integration of the perceived prominence patterns with the discourse structure.

In particular, I suggest that the first positive peak between 180-300 ms can be functionally interpreted as a P2. The early positivity, nonetheless, signals a more refined response to the merely noticing of an auditive modulation of pitch but it responds to the congruent condition. It appears, therefore, that L1 participants notice and identify the most appropriate pitch contour aligned to that type of target referent (part in whole-part relationships) in sentence-final position.

While confirming the N400 sensibility to pitch modulations in function to the degrees of activation, the appearance of this first positive peak represents new evidence emerging from this study.

If, in facts, German is a language structuring in its grammar the link between intonation and information status, the acquisition of German as native language warps the perceptual space of native speakers (Kuhl, 1995) making available a pitch accent phonological representation linked to the degree of activation of the referent, in this case, H+L* linked to accessible referents. The prolonged exposition to pitch variation during the experiment can cause therefore the subconscious recognition by participants of the underlying rule governing the experiment (Xu et al., 2006; Holt & Lotto, 2010). The recognition of patterns of association can trigger within the experiment a top-down activation of the relevant phonological representation, that is identified when encountered. The first P2 component associated with the congruent condition can, therefore, signal a categorical linguistic identification of a relevant phonetic pitch modulation that can be matched to an available native language phonological representation. Following the auditive matching, or non-matching, of an identifiable pitch contour, the N400 responds with a modulation of the following integration costs. A matching phonetic, phonological and pragmatic structure results in the easiest integration process, while non-matching prosody modulates the N400 negativity.

The higher cost remains, in a plastic language like German, the prosodic marking of completely given information, on a referent that is not textually given but forces the system to update model adding it to the available representation. Deaccentuation is therefore for native speakers the most misleading cue in the decoding of information structure.

The late positivity effect, as previously discussed, can be interacting with a positivity in response to high acoustical saliency of the L+H* condition and therefore is mainly anterior.

Interestingly, the late positivity is somewhat not aligned with the previous components. There is not a significant difference between Deaccentuation and L+H* in the very early time window, they both display a negative peak, and the N400 modulation signals that Deaccentuation is more difficult to integrate with respect to L+H*. Yet, the L+H* high peak is the condition with the largest late-positivity effect.
It is possible therefore that the auditive response in the early time window reflects an anticipatory process of the recognition of the congruent prosodic manipulation. The direction of the effects opposite and equal for both the incongruent conditions can be interpreted as the recognition that neither of the two is the congruent one, but the two incongruent patterns are fully disambiguated once the L+H* is encountered. After 400 ms, therefore, the L+H* is slightly rising and therefore similar to the original H* accent. The extremely high saliency and the contrastive instruction become available during the processing of the stressed syllable of the final noun. The different function of the accent is thus decoded later compared to the congruent PA, and also later than the Deaccentuation condition that is previously signaled by a peak on the preceding verb.
The evaluation of the presence of a focused unit in this condition might, therefore, require a deeper late cognitive activation since the whole discourse has to be entirely re-evaluated to identify the information that must be contrasted. The stronger positivity can be therefore an index of a pragmatic restructuring of discourse after a violation is encountered. In this respect, the paradigm as it is constructed does not allow to disambiguate in this condition between the effects due to acoustical saliency, to the narrow focus evaluation, and the processing of semi-active information; therefore, the component has to be considered as a reflection of the all merged factors. What is undoubtedly clear is that the positivity is different from the one found in the original Schumacher & Baumann (2010) experiment, and the difference is guided by the introduction of a narrow focus manipulation. Overall the results appear therefore in favor of what is proposed by Chen et. al., (2012) and Li et al., (2017) indicating that prosody can be independently modeled on the basis of referential Givenness and the presence of focus and that focused unit activate attentional response mechanisms to a higher extent compared to non-focused one.

4.10.3.2.2. L2

In L2, both early positivity in the 180-300 ms time window and the N400 modulation are two-way: congruent vs. incongruent. It is important to remember that on the basis of L1 systemic properties the reason why H+L* is a congruent condition is assumed not to be the matching with the referent's information status, but the matching to metrical structure of a broad focus sentence.
If, the early positive peak in the early time window reflects a recognizable top-down identification of a recognizable appropriate phonological structure then results in L2 are congruent with the explanation, and the positive peak signals the identification of an appropriate broad focus prosodic marking. Moreover, if, the N400 window is sensible to the appropriateness of an accent in function of the degree of cognitive activation of accessible referents, then L2 speakers' results highlight that the Deaccentuation contour and the L+H* contour increase the costs of discourse integration equally. The facilitatory effect is granted only by the H+L* condition.
Moreover, observing the ERP patterns in L1, it is possible to notice: that the modulation of the peak and respective positive polarity of the N400 component in the following window of the appropriate condition are somewhat similar, and that the N400 modulation is apparently larger than the first peak, especially in the right hemisphere areas. On the contrary, this relation seems to be inverted in L2 speakers. They show, in fact, a stronger peak in the early time window and a much smaller effect in the N400 window. Besides,

all 3 accents in the early time window have the same polarity, whether in L1 speakers the 2 incongruent accents have a negative polarity.
Interpreting the first positivity as P2, and phonological matching of appropriate pitch modulation, then the waveform in L2 appears to indicate that also the other 2 incongruent conditions receive a higher level of attention compared to native speakers, to a lesser extent compared to the matching H+L*. This can reflect a higher level of uncertainty and possibly the absence of an anticipation process. Reminding that the results are time-locked to the article at the beginning of the NP, it is possible that L2 speakers are slower in the integration of information coming from all 3 levels: prosody, lexical meaning, and discourse-pragmatic structure. It is possible that the full integration of discourse awaits the beginning of the stressed syllable of the target referent.
This delay could explain a rather smaller positivity effect in both early and N400 time window.
Overall this suggests that L2 speakers do not activate a top-down mechanism during the experiment in function of the degree of activation of the referent, but instead that they bottom-up await, perceive, and evaluate the congruency of information strongly relying on a bottom-up process.
In this perspective, the P2 signals that during the experiment they have learned that the sentences might prosodically end in different ways and they auditorily respond to the modulation at the time of the beginning of the NP, but that deviating prosody modulations are taken as anticipatory cues of the coming pitch modulation, and they are not linked to a 3-way discourse structure modulation, but only to a matching vs. non-matching broad focus accentuation pattern. It appears, therefore, that the presence of Deaccentuation does not significantly impact more than the processing of a highly salient peak during the integration of information.
Interestingly in the late positivity time-window, L2 perceivers show a high sensibility for the three accents modulation and the presence of L+H* shows the greatest effect.
Interpreting the positivity as in L1 speakers, therefore as a composite component comprehending: the increase processing effort necessary to access the information related to the final referent and the update of the mental representation of discourse, seeking in addition for information to be contrasted; and as well a higher attention allocation in response to acoustical saliency of a prominent F0 peak, L2 speakers show a comparable processing of native speakers.
The prominent L+H* peak causes the greatest effort and requires the higher level of attention allocation, followed by Deaccentuation and from the congruent condition.
The evidence emerging from the late-positivity window, considered together with the two-way modulation elicited in the two earlier time windows highlights that, the L2 speakers do not process the PAs on the basis of the cognitive activation of a referent, but are able, adopting a bottom-up strategy, guided by expectancies generated by the broad focus default structure, to evaluate the presence of an incongruent narrow contrastive focus. The F0 peak in the contrastive focus condition does not carry conflicting information to integrate the single referent in the discourse but carries conflicting information in the evaluation and update of the whole model. It appears from the results that L2 speakers were able to perceptually notice an incongruent modulation (two-way effect in early time window) but they did not consider it relevant in the accessing of meaning and integration of the referent (two-way N400 effect). In final instance, they evaluated what type of information necessitated to be added in the available representation of discourse. In this final stage, prominence is interpreted as a signal to allocate more attention to the specific referent, and therefore the whole discourse is deeply re-evaluated to match the contrastive information of the accent. The match is not found.

4.10.3.2.3. L1 vs. L2: The processing of Deaccentuation

L1 vs. L2 comparison of un-subtracted waves highlights a main effect of Deaccentuation in the early time-window and in the N400 window, more enhanced over right hemispheres electrodes.
In particular L1 speakers display larger negativities associated with Deaccentuation in these time windows and topographies, compared to L2 speakers.
Further comparisons highlight a main topographic processing difference for all 3 accents again in the early 180-300 ms time window, where all 3 accents display a positivity in L2 speakers.
Addressing under an SLA perspective the acquisition of the prosodic marking in L2, and considering the previous evidence reporting in perception and production studies the lack of the use of Deaccentuation to mark the givenness status of a referent, these results are strongly in line with the Markedness Differential Hypothesis and with the literature that highlights a structural resistance of the Italian intonation system to deaccent in sentence-final position.
L2 speakers show a main processing difference compared to L1 speakers that emerges in a key time-window, reflecting that in the on-line perception the auditory response to pitch manipulation in function of discourse structure is significantly different.
Namely, L2 speakers appear not to make use of an available consolidate linguistic strategy that associates Deaccentuation with fully given referents and evaluates all 3 accents on the basis of a bottom-up strategy.
The bottom-up evaluation, then, is strongly sensible to acoustical prominence and allows, therefore, to properly consider in the construction of discourse the presence of a recognizable prosodic manipulation, associable with contrastive information also in their L1, and its non-matching status with the already processed context.
From these results it follows that, the triggering of a comparative perceptual process between the properties of the target language and the properties of the native language, identified as necessary for the L2 acquisition of phonological and phonetics aspects by the available acquisitional models (Flege, 1995; Best & Tyler, 2005), has to be guided by attentional and control mechanisms. In this specific case, during L2 on-line processing, the association between a broad focus structure and H+L* has to be dismissed and on the basis of bottom-up evaluation of the stimuli, and with the acquired L2 competencies, the role of the semi-active level of activation of the referent has to be established as the main parameter for the correct integration of discourse. Only in that perspective deaccenting an accessible referent can be an incongruency, costlier than over-accenting an accessible semi-active referent (N400). The difference between the two accents nonetheless shows not to be considered in the early time window and not relevant in the N400 window. The pitch modulation, therefore, is considered strange in association with the content and syntactic structure of the stimuli, but the further detailed implication, related to the fact that the two pitch accents are in one case strangely low, and in the other one strangely high, is not attended and subsequently not used, in the processing of the single referent. Only in a later stage once a contrastive narrow focus, similar to a manipulation of focus available in L1, is processed and attended, the incongruency of this information contributes to the cognitive effort to add the referent to the model of discourse.
These results support the view that at the base of the two languages there is a different mapping of the relation between prosody and discourse and the despite the pool of participants is composed by intermediate and advance learners living in Germany at the time of the experiment, the relationship has not been fully acquired, and on-line processing is performed adopting the linking association that native speakers of Italian naturally use.
On one side, these results confirm that the association between prosody and information structure is more marked in this respect in German, compared to Italian. German speakers show to attend both the levels of

manipulation, the prosodic marking of different degrees of referential Givenness and the presence of a highly prominent peak narrowing the scope of focus. Italian speakers, on the other hand, show to evaluate the congruency of the discourse model mostly on the basis of focus properties, broad vs. narrow and do not use the lack of prosodic marking as an indication for fully active information. In this respect, the German system is more marked than the Italian one and confirming the MDH, the acquisition of a more marked phonological structure in adulthood appears not to be straightforward for L2 learners.

On the other side, if L2 learners preferably make use of the prosodic marking of broad and narrow focus, and systematically do not consider relevant a pitch modulation deaccenting a specific referent, the proposed model of the Effort Code, linking the perceptual prominence of prosodic marking in accordance with a degree of informativeness of the information conveyed in specific parts of the discourse, then shows to be in line with evidence reported by Chen et al., (2014) and Li et al., (2017). It modulates a universal principle guided by attentional mechanism strongly related to the properties of a prominent unit, acoustically salient. This mechanism can be guided in two independent and orthogonal ways marking the hierarchical organization of referential givenness of specific referents, and/or marking a communicative intention of the speaker, willing to highlight specific contents independently from their information status. L1 and L2 learners show a comparable sensitivity to the pitch contours within a relational perspective; therefore, both are correctly processing the marking of the broad or narrow focus. On the other hand, L2 learners lack the sensitivity and systematically do not allocate any attentional resource to Deaccentuation, and therefore they do not make use of its relationship with referential givenness and the modulation of pitch on the basis of the level of activation of a referent.

I propose, in final instance, that the specific link between intonation and referential givenness can be identified as the specifically marked relationship present between intonation and pragmatic discourse structure in West-Germanic languages, that it has never been reported in Italian.

4.12. Results: MFA and Cluster Analysis of Screening variables

In Chapter 1, paragraph 1.6., the role of external and internal factors on the L2 acquisitional success has been introduced. In the introduction and method of this Chapter, the variables investigating the external factors of the L2 acquisition of German of the group of L2 participants have been presented. In the current study, they have been collected through a sociolinguistic screening questionnaire. The further 2 questionnaires investigating the internal factors measured in this thesis: the Autism Spectrum Quotient and the Interpersonal Reactivity Index have also been introduced in the general introduction in Chapter 1, and adopted in the previous Chapters 2 and 3.
I proceed therefore performing a Multiple Factor Analysis (MFA) (J. Pagès 2002), firstly of the external factors impacting, and separately of the internal factors. The results of each MFA are taken as entry data for the construction of significantly diverging clusters within the sample population. The EEG signal of the newly formed clusters is then compared to observe a substantial difference in the modulation of the effects, with a particular interest in the analysis of the Deaccentuation condition. In final instance, the most relevant variables belonging to external and internal factors are selected to construct two major clusters of individuated language talented vs. non-talented learners, and the EEG waveform of these two final clusters will be compared and confronted with the results of L1 speakers. The aim is to show that the individuated talented learners show a modulation of the effect significantly similar to the modulation elicited by native speakers compared to the individuated non-talented group. The native-like prosody processing will show to be related to the presence of a specific combo of individual characteristics, dealing with the study of a language and with internal factors of the subject.

To perform the MFA and cluster analysis the FactoMineR and factoextra R packages are adopted (Kassambara & Mundt, 2016).
The screening questionnaire elicited 32 relevant variables: 9 are numerical (marked with the letter 's'), 23 are nominal (marked with the letter 'n').
The 32 variables are the following:
1) age of participants at the time of the experiment (Age, s); 2) gender (Gender, n); 3) macro-area of provenience in Italy (Origin, n); 4) competence in the local dialect (Dial_Comp, n): 5) use of local dialect on a daily basis in Italy (Dial_Use_Prod, n); 6) years of training of a musical instrument (Years_M_Inst, s); 7) years of specific training in singing (Years_Sing_Train, s); 8) self-declared level of Listening skills in English (ENG_List, n); 9) self-declared level of Interaction skills in English (ENG_Inter, n); 10) self-declared level of Oral Production skills in English (ENG_Oral_P, n); 11) self-declared level of Listening skills in German (GER_List, n); 12) self-declared level of Interaction skills in German (GER_Inter, n); 13) self-declared level of Oral Production skills in German (GER_Oral_P, n); 14) age of the beginning of the study of German as L2 (AoA, s); 15) age of the first immersion, of at list one month, in a German speaking area (Ao_1st_Immersion, s); 16) presence of a native speaker instructor in school classes of German L2 (School_L1_Teach, n); 17) presence of a native speaker instructor in free-time classes of German L2 (FT_L1_Teach, n); 18) presence of a native speaker instructor in University classes of German L2 (Uni_L1_Teach, n); 19) presence of specific training hours of spoken communication in German L2 in school classes (Speech_School, n); 20) presence of specific training hours of spoken communication in German L2 in free time classes (Speech_FT, n); 21) presence of specific training hours of spoken communication in German L2 in University Education (Speech_Uni, n); 22) total number of years of training in German L2 during school years (Years_L2_School, s); 23) duration of training in German L2 in free.-time classes (Time_FT_L2_Class, n); 24) total number of years of University education in German L2 (Years_Ger_L2_Uni, s) 25) spoken language used on a daily basis

during the immersion time in a German-speaking area (Spo_Lan_Imm, n); 26) duration of Immersion in a German-speaking area (time_Imm_ToEXP, n); 27) general occupation during the immersion in a German-speaking area (Occupation, n); 28) average number of hours per week of fruition of audio-visual content in German L2, such as TV and movies (TV_Hours, s); 29) average number of hours per week of listening to Radio in German L2 (Radio_Hours, s); 30) use of subtitles during the fruition of audio-visual material (Subtitles, n); 31) perceived perception and comprehension difficulty during the fruition of audio-visual material in German L2 (Diff_TV, n); 32) perceived perception and comprehension difficulty during the listening to Radio in German L2 (Diff_Radio, n);

To perform the MFA analysis, the 32 variables are organized in the following 14 groups (the variables in each group must be of the same type, either numerical or nominal; therefore, in the case of variables that measure a related construct but cannot be converted to the same numerical of factorial format without losing information, I created a group composed of one singular variable):

1) Age of Participants at the time of the experiment (A):
   composed of the single numeric variable Age;
2) Biographic information (Bio_Info):
   composed of the 4 nominal variables: Gender, Origin, Dial_Comp, Dial_Use_Prod;
3) Music Training (Music):
   composed of the 2 numerical variables: Years_M_Inst and Years_Sing_Train;
4) CEFR Levels pertaining speech in English (CEFR_ENG):
   composed of the 3 nominal variables: ENG_List, ENG_Inter, ENG_Oral_P;
5) CEFR Levels pertaining speech in German (CEFR_GER):
   composed of the 3 nominal variables: GER_List, GER_Inter, GER_Oral_P;
6) Age of Acquisition of German and Age of Arrival in a German speaking area (A_Ger_L2):
   composed of the 2 numerical variables: AoA, Ao_1st_Immmersion;
7) Presence of a native speaker instructor during the education period (L1_Instructor):
   composed of the 3 nominal variables: School_L1_Teach, FT_L1_Teach, Uni_L1_Teach;
8) Presence of specific training hours for oral production in German (L1_Speech_Training):
   composed of the 3 nominal variables: Speech_School, Speech_FT, Speech_Uni;
9) Length of Education in German L2 during school years (Time_L2_Sc):
   composed of the single numeric variable Years_L2_School;
10) Length of Education in German L2 in Free Time (Time_L2_FT):
    composed of the single nominal variable Time_FT_L2_Class;
11) Length of Education in German L2 during University (Time_L2_UNI):
    composed of the single numerical variable Years_Ger_L2_Uni;
12) Duration and socio-linguistic aspects of the Immersion time (Immersion):
    composed of the 3 nominal variables Spo_Lan_Imm, time_Imm_ToEXP, Occupation;
13) Hours per week of exposure to Oral Media in German (Hours):
    composed of the 2 numerical variables TV_Hours and Radio_Hours;
14) Perceived difficulty in comprehension of Oral Media in German and use of subtitles (Perception_Level):
    composed of the 3 numerical variables Subtitles, Diff_TV, Diff_Radio.

4.12.1. Results:

If the total amount of variance were partitioned in 32 equal dimensions, each dimension would explain a threshold level of 3.1 % of variance (3,1%). The first 11 dimensions, explaining a proportion of variance above the threshold level, sum up to 75 %, (see Figure 4.12.1). I consider in the following analysis the first 5 dimensions (Dim), explaining a total proportion of 47% of variance, and having an eigenvalue higher than 1 (Table 4.12.1).

	Eigenvalue	Explained variance	Cumulative variance
Dim 1	6.18	14.62	14.62
Dim 2	4.30	10.18	24.80
Dim 3	3.42	8.08	32.88
Dim 4	3.10	7.34	40.22
Dim 5	2.73	6.79	47.02

Table 4.12.1 Proportion of exp. Var. of the first 5 Dim

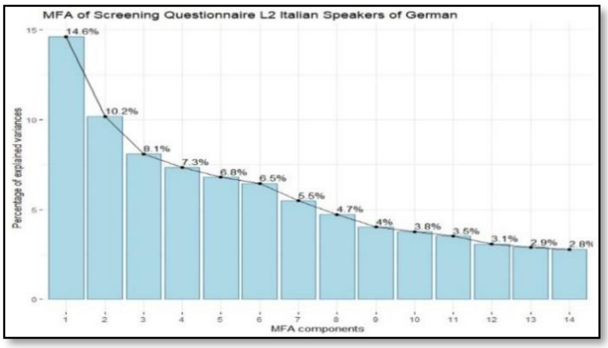


Figure 4.12.1 Explained variance of MFA's dimensions

The following Table 4.12.2. and Table 4.12.3. reports the contribution of each group to the dimension. The table is divided in 2 parts: from Dim1 to Dim 3 and separately reports Dim4 and Dim5.

Groups	Dim1	Groups	Dim2	Groups	Dim3
L1_Instructor	13.26	Time_L2_FT	16.04	CEFR_GER	17.10
L1_Speech Training	13.25	Immersion	14.75	Time_L2_FT	13.25
Time_L2_FT	12.77	L1_Instructor	14.61	Immersion	11.41
Time_L2_Uni	9.44	Bio_Info	12.29	Bio_Info	9.22
Immersion	8.42	CEFR_GER	10.35	L1_Instructor	9.22
CEFR_ENG	7.30	Perception_Level	9.30	CEFR_ENG	9.18
A_GER_L2	7.28	L1_Speech_Training	8.02	Perception_Level	7.73
Bio_Info	5.44	CEFR_ENG	6.57	L1_Speech_Training	6.57
A	5.38	A_GER_L2	3.24	A_GER_L2	5.51
Perception_Level	4.66	Music	1.86	Music	5.32
Time_L2_Sc	4.07	Hours	1.80	Time_L2_Sc	2.49
Music	3.54	A	0.54	Hours	1.42
CEFR_GER	2.64	Time_L2_Sc	0.51	A	1.33
Hours	2.54	Time_L2_Uni	0.13	Time_L2_Uni	0.26

Table 4.12.2 Contribution of each group to the first 5 dimensions (Dim 1 – Dim 3)

Groups	Dim4	Groups	Dim5
Time_L2_FT	17.14	CEFR_GER	15.11
CEFR_ENG	16.58	L1_Speech_Training	12.61
Immersion	11.98	Hours	11.14
L1_Instructor	10.98	L1_Instructor	10.91
L1_Speech_Training	8.88	Perception_Level	10.70
A	7.52	Immersion	9.56
Bio_Info	5.52	Bio_Info	8.87
CEFR_GER	4.53	Time_L2_FT	7.05
A_GER_L2	4.33	Time_L2_Sc	5.74

Hours	3.63	CEFR_ENG	2.96
Music	3.24	A_GER_L2	2.69
Perception_Level	3.23	Time_L2_Uni	2.60
Time_L2_Uni	2.02	Music	0.06
Time_L2_Sc	0.41	A	0.00

Table 4.12.3 Contribution of each group to the first 5 dimensions (Dim 4 & Dim 5)

The following Table 4.12.4. and Table 4.12.5 report how well each group is represented in the dimension (squared cosine). A very well represented group, when the interaction with a second dimension is considered, has squared cosine values that sums up to a value close to 1.0.

Groups	Dim1	Groups	Dim2	Groups	Dim3
Time_L2_UNI	0.34	Time_L2_FT	0.12	CEFR_GER	0.13
L1_Speech_Training	0.24	Immersion	0.12	Time_L2_FT	0.05
L1_Instructor	0.19	L1_Instructor	0.11	Immersion	0.04
A_GER_L2	0.18	Bio_Info	0.09	CEFR_ENG	0.04
Time_L2_FT	0.16	CEFR_GER	0.07	A_GER_L2	0.03
A	0.11	Perception_Level	0.06	Bio_Info	0.03
Immersion	0.08	L1_Speech_Training	0.04	L1_Instructor	0.03
CEFR_ENG	0.08	CEFR_ENG	0.03	Perception_Level	0.03
Time_L2_Sc	0.06	A_GER_L2	0.02	Music	0.02
Bio_Info	0.03	Hours	0.01	L1_Speech_Training	0.02
Perception_Level	0.03	Music	0.00	Time_L2_Sc	0.01
Music	0.03	A	0.00	Hours	0.00
Hours	0.02	Time_L2_Sc	0.00	A	0.00
CEFR_GER	0.01	Time_L2_UNI	0.00	Time_L2_UNI	0.00

Table 4.12.4 Squared Cosine of each group to the first 5 dimensions (Dim 1 – Dim 3)

Groups	Dim4	Groups	Dim5
CEFR_ENG	0.10	Hours	0.09
Time_L2_FT	0.07	CEFR_GER	0.07
A	0.05	L1_Speech_Training	0.05
Immersion	0.04	Perception_Level	0.04
L1_Instructor	0.03	L1_Instructor	0.03
L1_Speech_Training	0.03	Time_L2_Sc	0.03
A_GER_L2	0.02	Immersion	0.02
Hours	0.01	Bio_Info	0.02
Bio_Info	0.01	Time_L2_FT	0.01
CEFR_GER	0.01	Time_L2_UNI	0.01
Music	0.01	A_GER_L2	0.01
Time_L2_UNI	0.00	CEFR_ENG	0.00
Perception_Level	0.00	Music	0.00
Time_L2_Sc	0.00	A	0.00

Table 4.12.5 Squared Cosine of each group to the first 5 dimensions (Dim 4 & Dim 5)

Figure 4.12.2 reports the groups distributions in the sample, according to group coordinates, selecting Dim1 and Dim5. Despite Dim2 explains the second highest proportion of variance and has the second highest values of group contributions in the dimensions, I report here the plot representing the matching the first and fifth dimension (Dim1 and Dim5). Dim5 assigns a greater weight to different variables compared to Dim1 (e.g. Hours), allowing to discriminate participants along the dimension adopting a fuller spectrum of the available variables and therefore allowing to appreciate different characteristics of the sample.

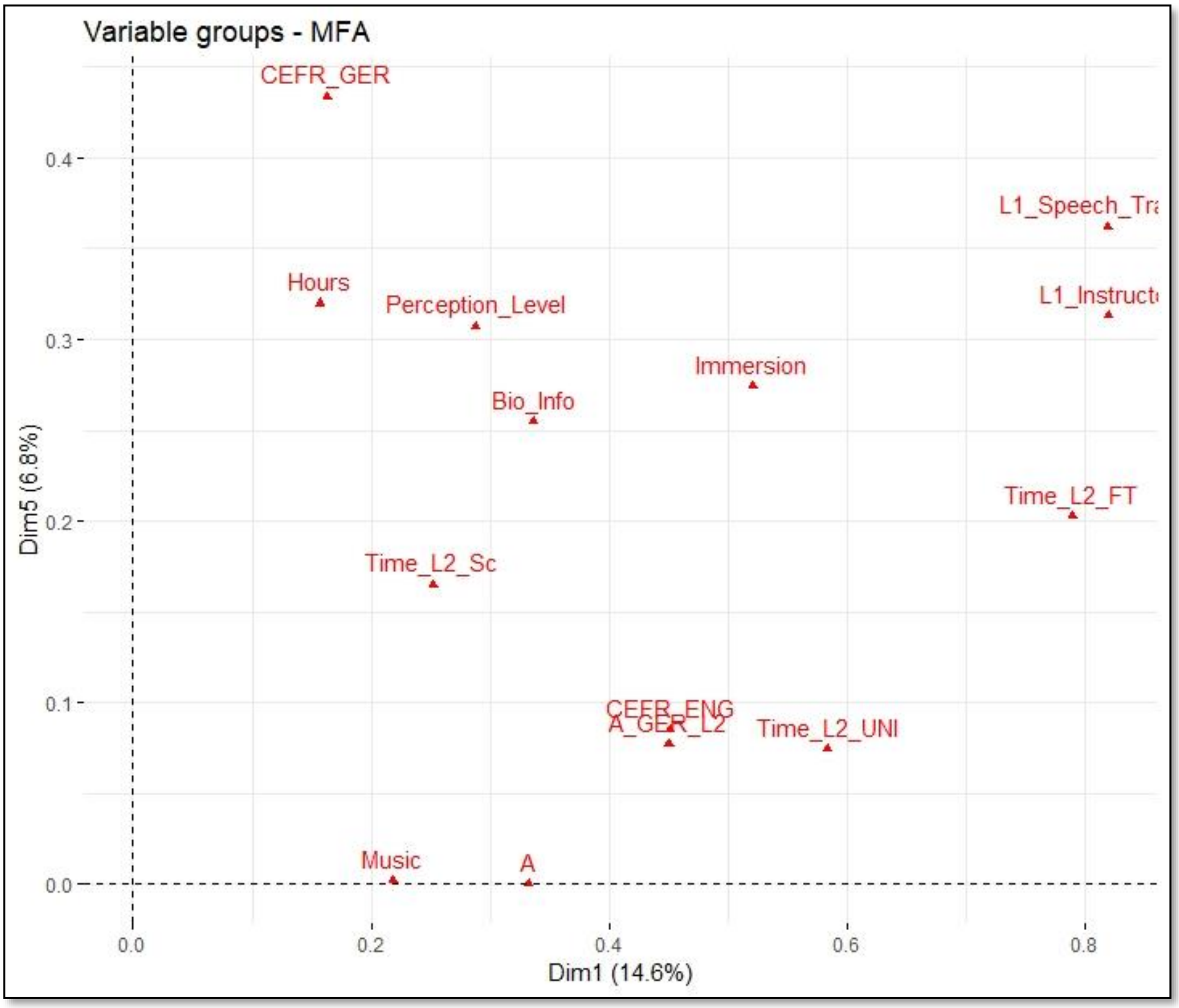


Figure 4.12.2 Groups distributions in the sample, according to group coordinates, selecting Dim1 and Dim5

From the previous plot it is possible to see that Dim1 organizes the groups dividing the sample between individuals well represented in the variables CEFR_GER (self-declared CEFR level in oral interaction skills in German as L2), Hours (self-declared hours of exposure to audio-visual material in German) and Perception_Level (how difficult the individual perceive it is to watch and listen to audio-visual material in German L2) on the left part of the plot, and individuals, on the right side of the plot, well discriminated by the variables belonging to the group L1_Speech_Training (the presence of hours of training devoted to only spoken languages in school, free time and university education), L1_Instructor (the presence of native speakers instructors in school, free time and university education), and Time_L2_FT which measures the duration of free time classes' attendance alone.

Dim5 organizes the sample in individuals discriminated by the groups CEFR_GER, L1_Speech_Training, Hours, Perception_Level, and L1_Instructor on the top of the plot, and individuals discriminated by the group Music (years of musical education), A (age of participants), CEFR_ENG (self-declared CEFR level in oral interaction skills in English as L2), A_GER_L2 (age of acquisition of German and age of 1st immersion), and Time_L2_Uni (Years of University education devoted only to the acquisition German as L2), in the bottom of the plot.

It is possible to observe how single individuals are positioned in the sample, in each single variable, according to the most relevant groups for the sample discrimination.

I report here single plots (Figures from: 4.12.4. to 4.12.10) relevant both in Dim1 and Dim5: CEFR_GER, Hours, L1_Speech_Training, L1_Instructor .

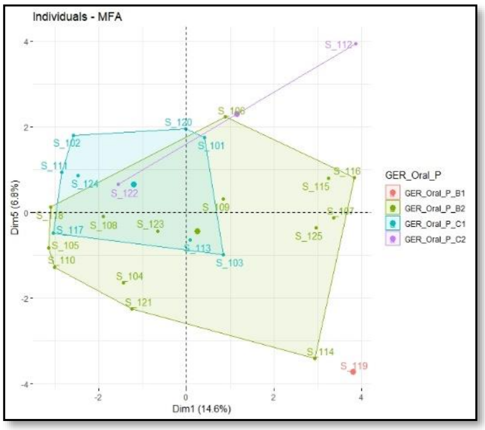


Figure 4.12.4 CEFR_GER_Listening

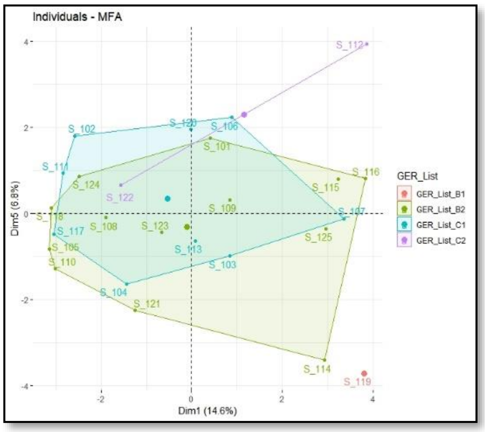


Figure 4.12.3 CEFR_GER_Oral_P

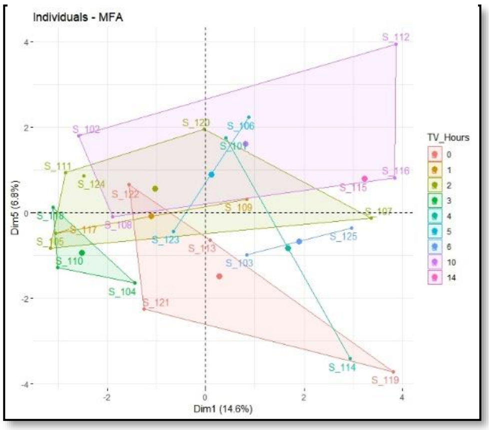


Figure 4.12.5 Tv_Hours

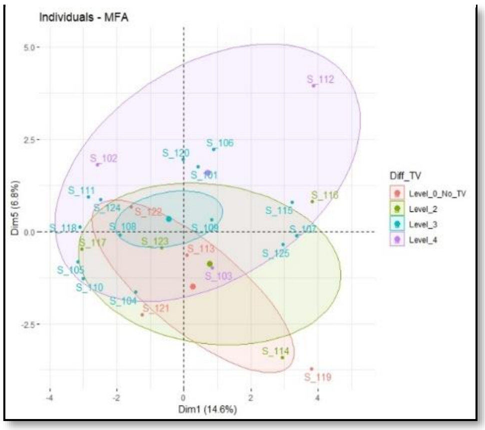


Figure 4.12.6 Diff_TV (higher level – greater comprehension)

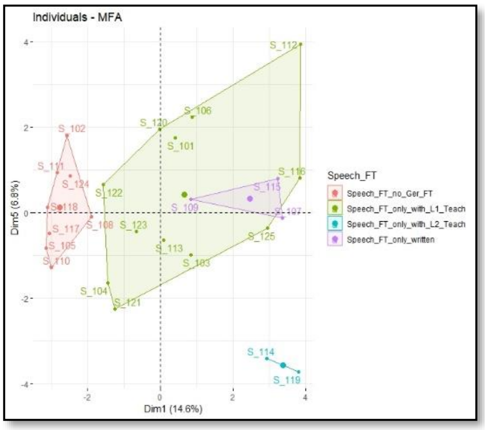


Figure 4.12.8Speech_Training in Free Time classes

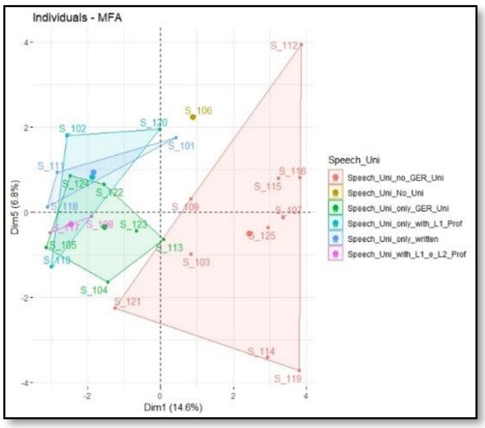


Figure 4.12.7 Speech_Training in University education

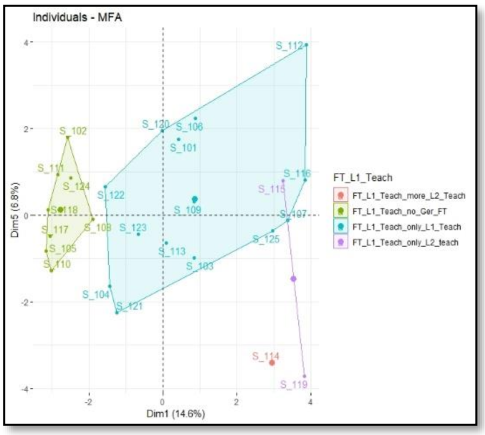


Figure 4.12.10 Presence of a L1 instructor in Free Time classes

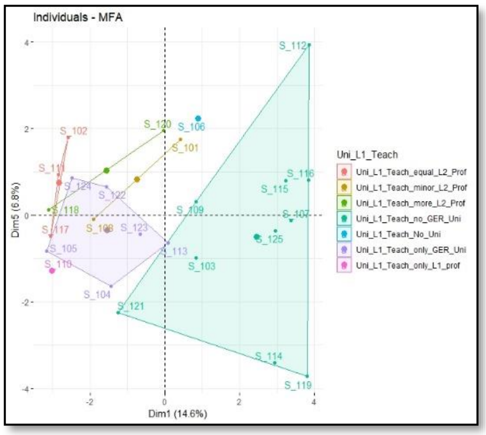


Figure 4.12.9 Presence of a L1 instructor in University education

In addition, I report the plots showing specific characteristic of the time of immersion, in time and socio-linguistical aspects (Figures from 4.12.12 to 4.12.14)

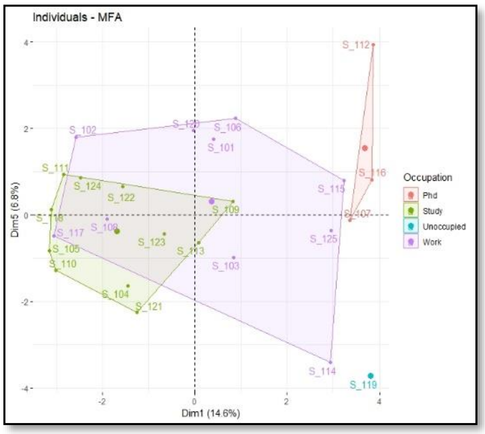


Figure 4.12.12 Occupation during immersion

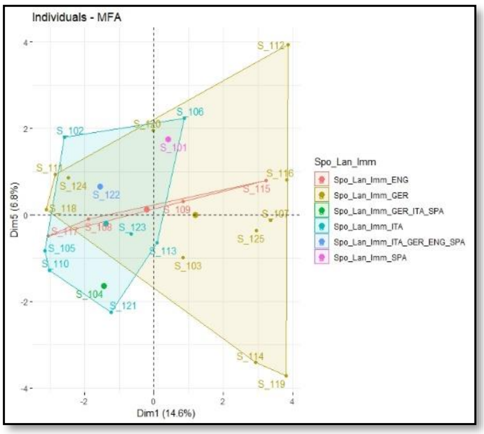


Figure 4.12.11 Spoken_Language during immersion

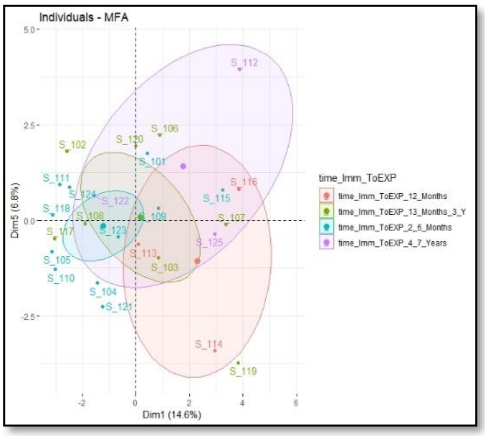


Figure 4.12.13 Immersion length at the time of the experiment

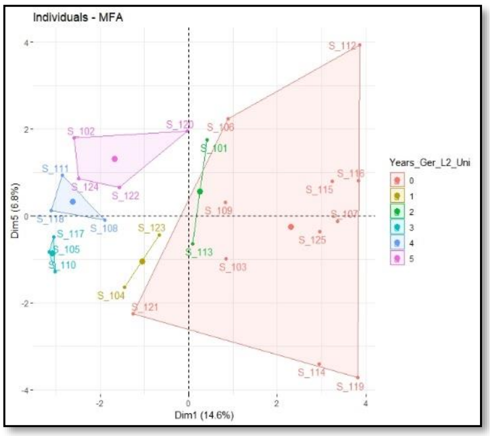


Figure 4.12.14 Years of University education specific for German L2

Following Dim1, it is possible to observe that the sample is firstly averagely organized in two macro subparts. One subpart is composed of learners that are currently graduating or have graduated pursuing a university education centered on the acquisition of German (and possibly a second L2) as a second language. This subpart has been in living, at the time of the experiment, for a rather short time (when compared to the rest of the sample) in a German-speaking area.
The second subpart, on the right side of the plot, is composed by a population who has not specialized during university education specifically in the acquisition of German, but that reports longer length of immersion, which is possibly the reason for the acquisition of German as L2. They in fact report to be living in a German area for work or higher education reasons. The first subpart (left side of the plot) is further dividable in individuals who are graduating in a German-speaking area, and individuals that are graduating in another area (most probably Italy). The formers, therefore, are required to speak the language in order to graduate successfully. The second subpart, on the other hand, also appears to train in the free time in specific classes of German L2, apart from their education. One relevant observation to make is that classes in the free time, probably taken while already in immersion, are overall conducted by native speaker instructors, while individuals formally receiving an education outside a German-speaking area, are exposed to the spoken language of both native and non-native instructors. It appears, therefore, that the length of training and education is higher for the subpart graduating in fields specifically targeting the acquisition of German as L2, while the population acquiring German for work and life-related reasons is consistently exposed to the oral language spoken by native speakers to a greater extent, when compared to the first subpart. The higher exposure is also captured by the group of variable Hours, which in the case of Tv_Hours is less clear-cut compared to variables distributed along Dim1 but shows that individuals in the top part of the plot are consistently more exposed to audio-visual material compared to individuals on the bottom part of the plot. Individuals on the top-right area are, therefore, individuals who have learned German for work-related reasons, that at the time of the experiment have been living for the most prolonged period of time in a German-speaking area and have been more exposed than other to the spoken language of native speakers.

4.12.2. Cluster Formation

Results of the MFA analysis are taken as input for Hierarchical Clustering on Principal Components (HCPC). The aim is to observe whether different clusters of the sample, on the basis of the screening properties that have been collected, pertaining the type of specific education on German as L2 and properties of the time of immersion of the individuals in the sample, can account for processing differences during the perception of specific pitch contours in German L2. In particular, whether a longer specific type of education or more extended immersion's length has an impact on the processing of intonational contours guided by the integration the discourse structure, and specifically related to the processing of the bridging inferences created across the context and target sentence. In particular, we have seen in the L2 analysis that L2 speakers mainly differ from L1 speakers in the processing of the Deaccentuation contour in the early time window and N400 time window. I exploratively observe, therefore, whether the clusters of population are related with the variation of the Deaccentuation's early and medial peaks, that have a positive direction in L2, and a negative direction in L1.

The results of the HCPC provide the following clusters' subdivision.

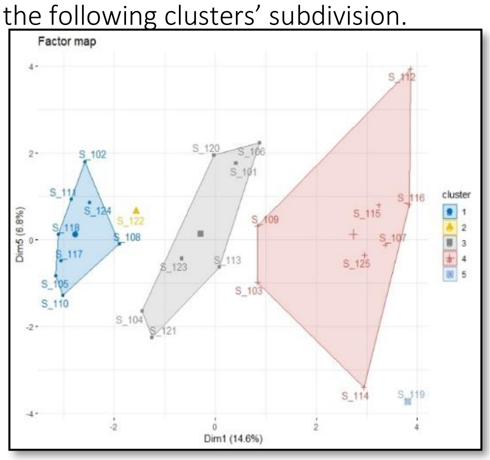


Figure 4.12.15 Cluster formation on the basis of the Screening Questionnaire

In the following Table 4.12.6. I report for each cluster, its composition and the most represented dimensions.

Clusters	Individuals	Dimensions
cl1	S_102 S_105 S_108 S_110 S_111 S_117 S_118 S_124	Dim2 Dim1
cl2	S_122	Dim2
cl3	S_101 S_104 S_106 S_113 S_120 S_121 S_123	Dim4
cl4	S_103 S_107 S_109 S_112 S_114 S_115 S_116 S_125	Dim1 Dim3
cl5	S_119	Dim3 Dim5

Table 4.12.6 Clusters on the basis of Screening variables

Individuals in cl1 averagely began to study German L2 at the age of 15, during school years, and averagely arrived in a German-speaking area at the age of 23. 7 individuals are female, originally from the south and main islands of Italy, 1 is male coming from the north. They do not report a particular training playing a musical instrument, two of them report 5 and 7 years of singing training. They homogeneously have a C1 self-declared CEFR level in English as L2, and they report B2 and C1 levels in German L2. None of them attended specific classes of German L2 apart from their education. 6 individuals are attending or attended the university education outside of a German-speaking area (probably, in Italy) and have been trained by both native and L2 instructors during their education; 2 individuals are graduating in a German-speaking area. They averagely have studied German during their school education for 3.2 years and have a university education in German L2 of 3.8 years. Therefore, they have an average training of 6 years that has begun during the last years of high school. At the time of the experiment, they had been living in a German-speaking area for a minimum period going from 2 months up to 3 years. They are at the time of the experiment studying or working in a German-speaking area. They are averagely exposed for 4 hours per week to audio-visual material in German and averagely declare a level 3 (out of 5) in the comprehension question, that corresponds to the relative ease in the fruition of audio-visual material. They can follow the content of TV and movies in German for a prolonged period of time, and they find it a recreational activity, but they report difficulties in the comprehension of jokes, implicit language, irony and sarcasm, and culture-specific topics, and are not able to recognize regional accents.

Individuals in cl3 have an average age of 23.8. 5 of them are females coming from areas spanning from the north to the south of Italy, 2 of them are males coming from the north. One of them has 6 years of musical training playing an instrument, none of them has specific training in singing. They report CEFR levels in both English and German spanning from B2 to C1, tendentially higher in German than English. They averagely began the study of German at the age of 16.6 and averagely arrived in a German-speaking area at the age of 21. Only two of them attended German classes during school years; nonetheless, these two averagely began during middle school. All of them attended specific classes aside from their education, where a native speaker instructor was present. Their University education is heterogeneous. The cluster is composed of both: 1 individual without a university education, 3 are graduating in a German-speaking area, 2 are graduating with a major in German outside a German-speaking area. At the time of the experiment, they averagely have a university training in German of about 1.6 years and their immersion-time spans as in cl1 from below 6 months up to 3 years. They report an exposure to both audio-visual material and Radio of about 2.7 hours each per week (summing up to more than 5 hours per week). As in cl1 they averagely report a level 3 in the comprehension question about audio-visual material. The main difference between cl1 and cl3 is, therefore, the greater exposure to native language speakers in L2 dedicated classes, present in cl3 and missing in cl1.

Individuals in cl4 have an average age of 27.8. 6 are females coming from areas spanning from the north to the south of Italy, 2 are males coming from the north. They averagely report a musical instrument training of 4 years. The reported CEFR levels pertaining English shows that half of the group has an average level of B2, and the remaining participants declare the highest C2 level; on the contrary, they averagely declare a lower level in German, mostly B2, except for one participant reporting a C2 level also in German. They averagely began to acquire German at the age of 21.5 and averagely arrived in Germany at the age of 25.4. Only two of them attended German classes during school, none of them is currently graduating or has graduated using German as communication-language or majoring in the acquisition of German as L2. They all have attended specific classes of German L2 apart from their education, 7 of them only with native-speaker instructors, only 1 with an L2 instructor. At the time of the experiment they had been living in a German-speaking area for a period of time averagely shifting from 1 year to 4. 3 of them have a higher Ph.D. education, 4 of them are in a German-speaking area for work reasons, only 1 is studying as an undergrad. They report a fruition of audio-visual material of about 6 hours per week, and 3.5 hours per week listening to Radio. In the comprehension question pertaining to the fruition of audio-visual material, their answers span from level 2 to level 4, showing higher variability when compared to the other clusters. The comprehension of spoken language used in Radio is declared by half of the individuals in the cluster as "rather easy" and from the remaining individuals as "rather hard"(out of 4 levels going from extremely easy to extremely hard).
This cluster mainly differs from the other two by including individuals that are older and arrived later in life in a German-speaking area for work reasons. None of them as a specific formation in German, but all of them received a specific intensive training with a native speaker instructor. The reduced training time appears to be compensated by a higher time of exposure to audio-visual material.

The remaining two clusters (cl2 and cl5) include 2 single individuals that display uncommon characteristic in their biographic information within the selected sample of population.
Cl2 corresponding to the individual S_122. She is a female speaker of the age of 26 coming from different areas in the north-center of Italy. She does not have any particular music education and declares C2 levels in both English and German L2. She began the acquisition of German at the age of 14 and graduated in a German-speaking area; therefore, she has a comprehensive training of about 10 years. She reports the exposure to native speaker instructors since school years and also in specific classes apart from her education. At the time of the experiment, she had been living in a German-speaking area for more than 4 years. Nonetheless, she declares no hours of exposure neither to audio-visual material nor to Radio.
Cl5, corresponding to the individual S_119, is a 33 years old male, coming from the islands-regions in Italy, and declares an average B2 level in English as L2, and B1 level in German as L2. He did not attend any German specific classes during school and University. He attended a specific L2 class apart from his education, but not with a native speaker instructor. At the time of the experiment, he had been living in a German-speaking area for a period ranging from 1 to 3 years.
Cl2 and Cl5 represent therefore the lower and higher tail of the sample's distribution, corresponding to individuals having either a relative lower competence in German L2 with respect to the population in the sample (cl5) and, on the opposite, a relative higher competence (cl2).

4.13. ERP – Deaccentuation wave in function of cluster's structure.

I plot and analyze in this section the modulation of the EEG waveform in the 3 accent-conditions in each cluster, investigating whether the individual differences across clusters can explain variability in the processing of the incongruent alignment of a Deaccentuation pattern and overtly acoustically salient contrastive focus peak, when compared to the processing of the congruent pattern for both broad focus sentences and accessible information. The EEG waveform for each accent is plotted to maintain the same association between colors and conditions as before, dashed and dotted lines are used to mark the cluster analysis. Cl2 and Cl5 are not plotted since they are composed of only one single outlier individual. I expect that the most evident difference in the size of the effects occurs between cl1 and cl4. An indicator of a processing demonstrating the acquisition in the direction of the native-like prosody perception is the three-way modulation of the 3 pitch accents, especially in the early time-windows. In particular in the N400 window, since the 3-way modulation in L1 speakers has been reported in Schumacher & Baumann (2010) and fully replicated in the previous analysis of this study. In addition, it can be interesting to observe whether a subpopulation with, in this case, a more extended training and a earlier Age of Acquisition and of Arrival (Age of 1st immersion), or on the contrary, a shorter but more intensive training and later Age of Acquisition and Age of Arrival, but longer duration of immersion, can modulate the auditive response in the first time window and the integration of information as well as the attention allocated to the contrastive peak in the late-positivity window.

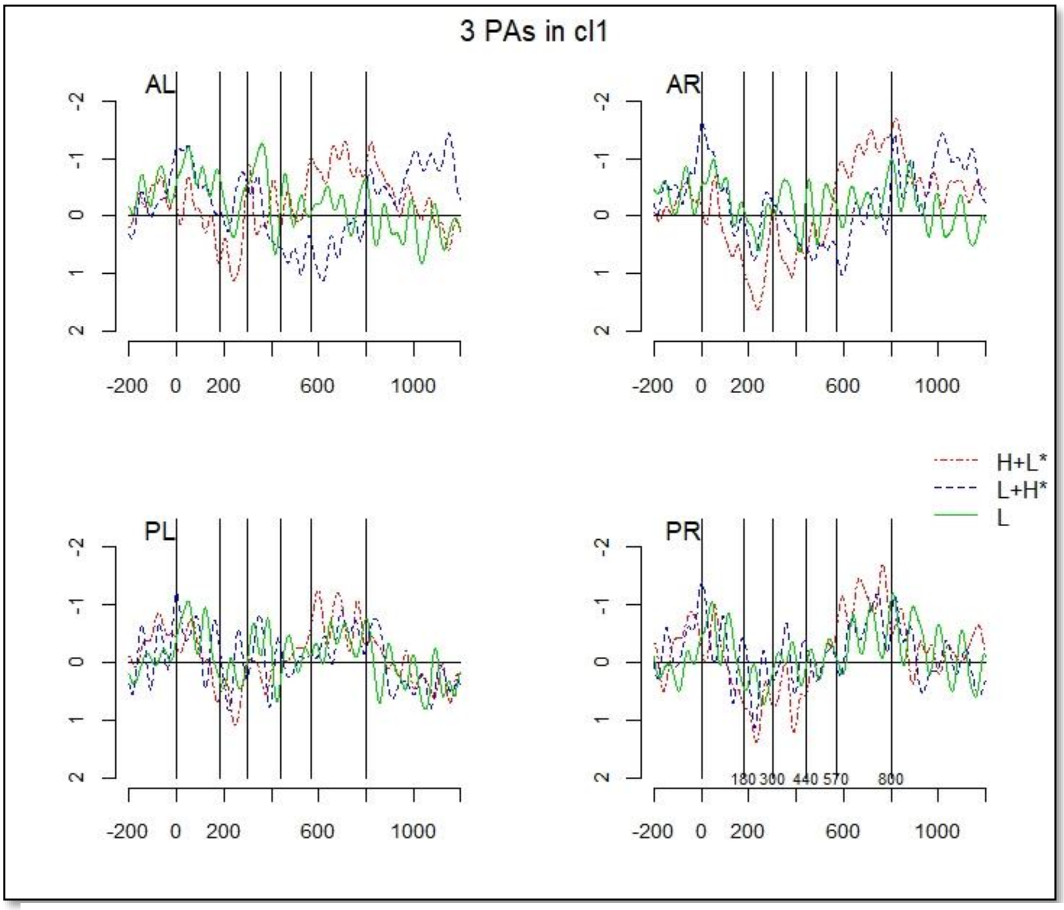


Figure 4.13.1 4 ROI - Cluster 1

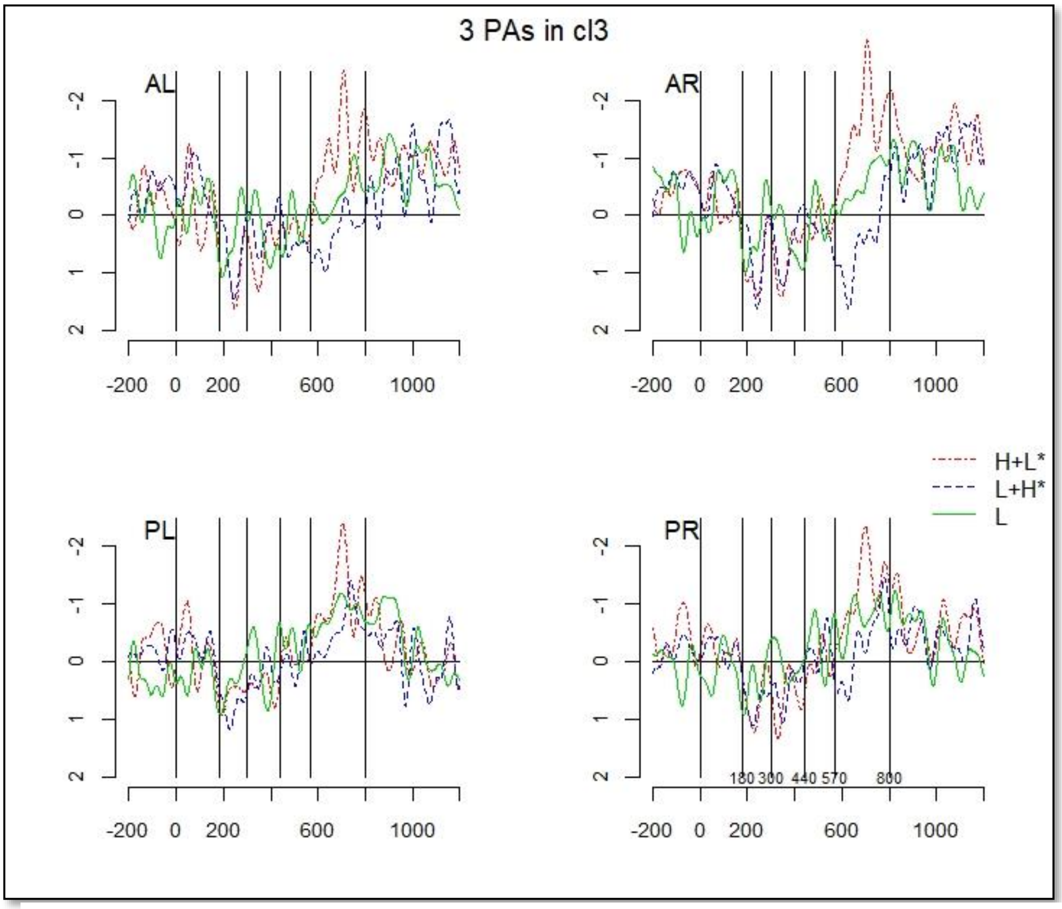


Figure 4.13.2 4 ROI - Cluster 3

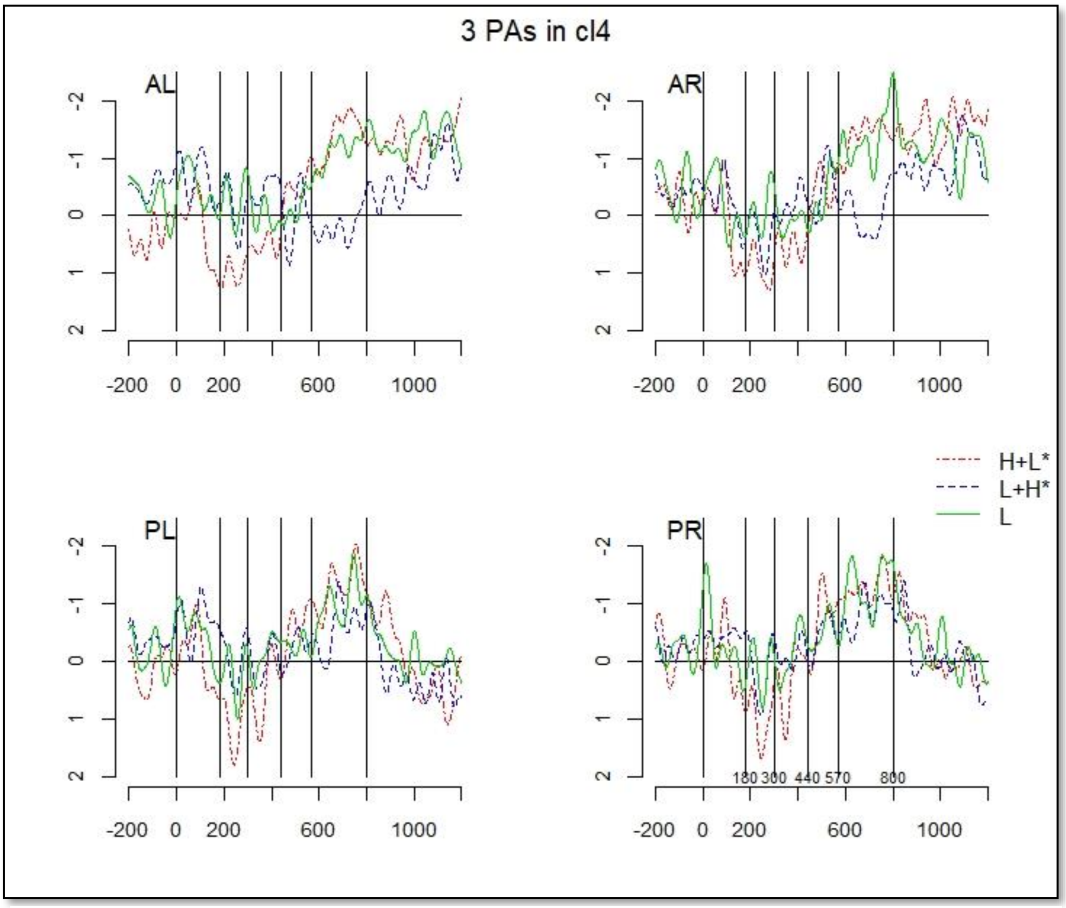


Figure 4.13.3 4 ROI - Cluster 4

4.13.1. Results of main factor PROSODY and lateral topographic distribution

4.13.1.1. Cl1 - Lateral

Over lateral electrodes, in the 180 – 300 ms time window, the data show the presence of tendency for PROSODY factor F(2,14) = 3.13 ; p = 0.08. In the 300 – 440 ms time window, there is no significant effect.
In the 570 – 800 ms time window, the data show an effect for Prosody F (2,14) = 5.52 ; p = 0.017 ; in addition, and an interaction of the topographic factor with the main prosody effect F(6,42) = 2.49; p = 0.04.

4.13.1.2. Cl3 - Lateral

Over lateral electrodes, in the 180 – 300 ms time window, the results show a tendency for PROSODY F(2,12) = 3.35 ; p = 0.07. In the 300 – 440 ms time window, there is no significant effect as in the previous cluster.
In the 570 – 800 ms time window, there is both an effect for Prosody F (2,12) = 21.53 ; p = 1.07 e-04 ; in addition, the results show a topographic effect of the Lateral factor F (3, 18) = 5.79; p = 5.95 e-03; and the presence of an interaction with the main prosody effect F(6,36) = 6.72; p = 7.82 e-05.

4.13.1.3. Cl4- Lateral

Over lateral electrodes, in the 180 – 300 ms time window, the factor PROSODY is significant F(2,14) = 7.82; p = 0.005. In the 300 – 440 ms time window, there is a tendency of the factor Prosody F(2,14) = 2.88 ; p =0.09; and in the 570 – 800 ms time window, there is both an effect for Prosody F (2,14) = 8.88 ; p = 0.003; an interaction of the topographic factor with the main prosody effect F(6,42) = 6.00; p = 0.0001.

4.13.2. Pairwise Comparisons in 4 ROI for each cluster

Pairwise comparisons, FDR corrected for 3 within-subjects contrasts, over the lateral factor, show significant differences in cl1 and cl3 only in anterior regions within the late-positivity window (570-800 ms).
In the Anterior Left quarter, cl1 shows a significant effect of PROSODY in all 3 comparisons: Congruent vs. Deaccentuation t (7) = 3.89; p = 0.009; C.I.(-1.10 - -0.27); Congruent vs. Contrastive t (7) = -5.57; p = 0.003; C.I.(-1.97 - -0.80); and Contrastive vs. Deaccentuation t (7) = -2.80 ; p = 0.03; C.I. (-1.29 - -0.11);
In the Anterior Right quarter, cl1 shows a significant effect of PROSODY in only 1 comparison: Congruent vs. Contrastive t (7) = -3.66; p = 0.02; C.I.(-2.05 - -0.44);
In the Anterior Left quarter, cl3 shows a significant effect of PROSODY in 2 comparisons: Congruent vs. Deaccentuation t (6) = -2.75 ; p = 0.05; C.I.(-1.44 - -0.08); Congruent vs. Contrastive t (6) = -4.71; p = 0.01 ; C.I.(-2.21 - -0.70);
In the Anterior Right quarter, cl3 shows a significant effect of PROSODY in all 3 comparisons: Congruent vs. Deaccentuation t (6) = -3.20; p = 0.02 ; C.I.(-1.71 - -0.23); Congruent vs. Contrastive t (6) = -7.13 ; p = 0.001; C.I.(-2.71 - -1.33); and Contrastive vs. Deaccentuation t (6) = -3.20; p = 0.02; C.I. (-1.83 - -0.25);
In cl4, pairwise comparisons, FDR corrected for 3 within-subjects contrasts over the lateral factor, show significant differences in the early and late components. Early components are broadly distributed, excluding the Anterior Right quarter, the late component is frontally distributed.

In the 180 – 300 ms time window, in the Anterior Left quarter an effect of PROSODY is found in one comparison: Congruent vs. Deaccentuation t (7) = 4.14 ; p = 0.01 ; C.I.(0.55 – 2.00); and a tendency in a second one: Congruent vs. Contrastive t (7) = 2.37 ; p = 0.08; C.I.(0.00 – 2.30). In the Posterior Left quarter, the data show a tendency for the same 2 comparisons: Congruent vs. Deaccentuation t (7) = 2.60; p = 0.07; C.I.(0.07 – 1.60); Congruent vs. Contrastive t (7) = 2.41; p = 0.07; C.I.(0.02 – 2.16). In the Posterior Right quarter, again two tendencies emerge for the same 2 comparisons: Congruent vs. Deaccentuation t (7) = 2.50 ; p = 0.06; C.I.(0.05 – 1.82); Congruent vs. Contrastive t (7) = 3.02; p = 0.06; C.I.(0.18 – 1.50).
In the 570 – 800 ms time window, in the Anterior Left quarter, an effect of PROSODY is found in two comparisons: Congruent vs. Contrastive t (7) = -5.79; p = 0.002 ; C.I.(-2.24 - -0.94) and Contrastive vs. Deaccentuation t (7) = -3.35; p = 0.02 ; C.I. (-2.29 - -0.39). In the Anterior Right quarter, the same significant effect is found in the same 2 comparisons: Congruent vs. Contrastive t (7) = -2.80; p = 0.04; C.I.(-2.52 - -0.21) and Contrastive vs. Deaccentuation t (7) = -3.52 ; p = 0.03; C.I. (-2.16 - -0.43).

4.13.3. Between clusters comparison of each condition

In this section, I investigate whether the cluster structure can further explain the variability in the L2 data, mainly focusing on the modulation of the ERPs in function of Deaccentuation.

I include in the comparison cl1, cl3, and cl4 and exclude from the analysis the remaining cl2 and cl5 because of the unbalanced number of participants. Then, I compare the EEG waveform of un-subtracted waves in each condition and between clusters.

Figures 4.13.4, 4.13.5 and 4.13.6. below show 3 plots, 1 for each accent, of the sample's cluster composition. In each of the 3 panels, cl1 has always the color of the plotted accent condition (red-appropriate; blue – contrastive; green-Deaccentuation) and cl3 and cl4 are respectively always plotted in purple and in black, to allow for a higher readability.

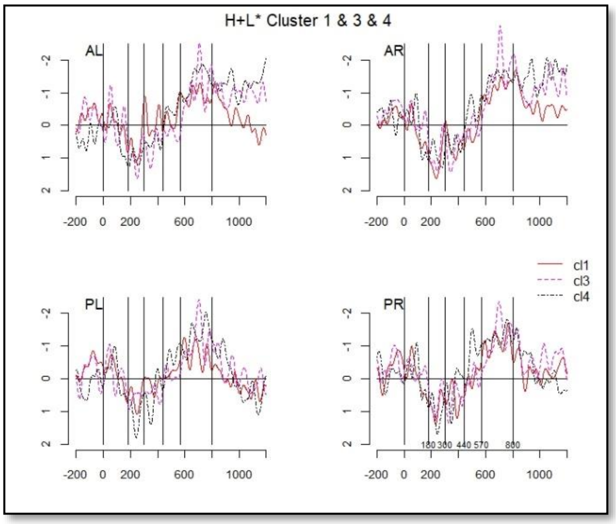


Figure 4.13.4 Congruent condition un-subtracted waves of cl1 cl3 & cl4

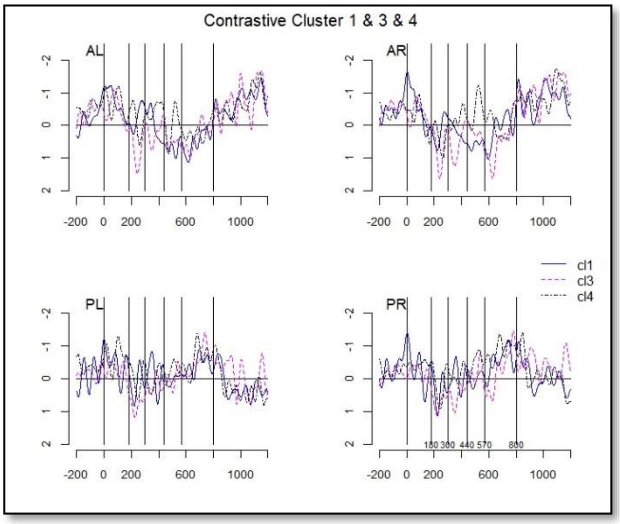


Figure 4.13.5 Figure 4.20.4 Contrastive condition un-subtracted waves of cl1 cl3 & cl4

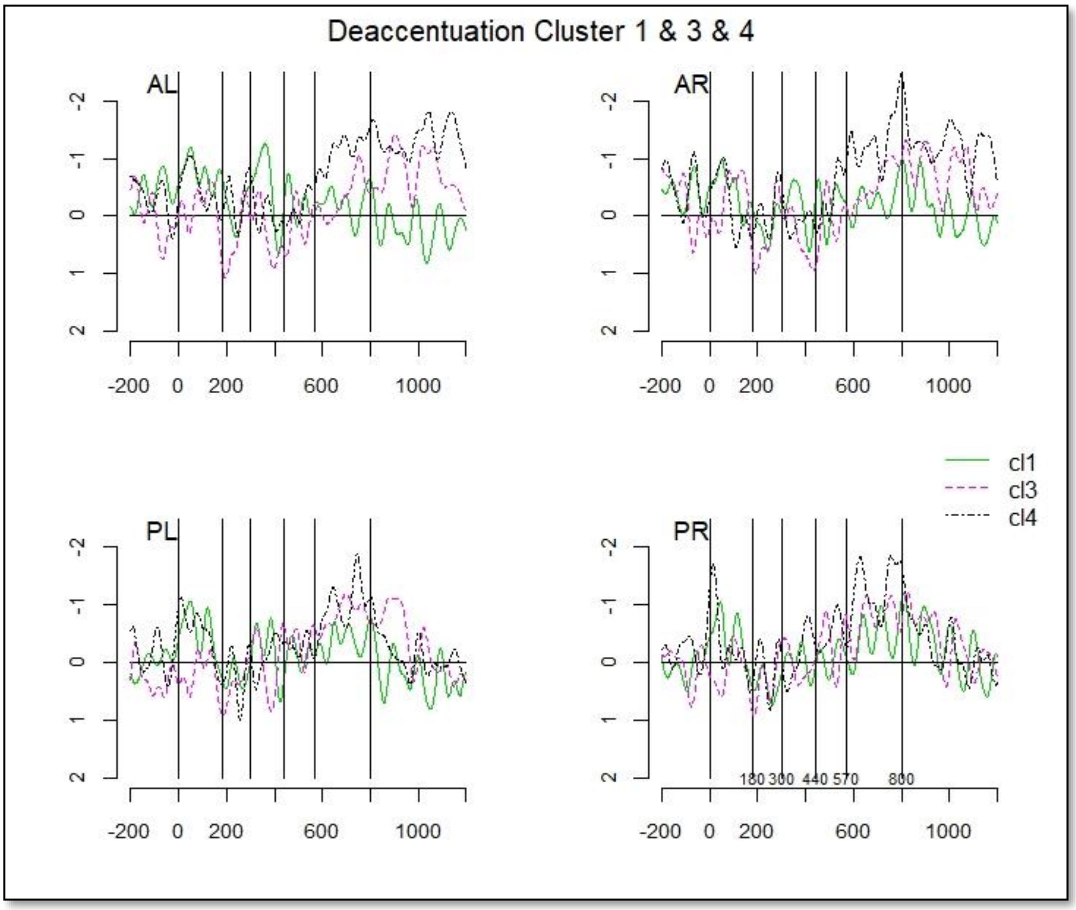


Figure 4.13.6 Figure 4.20.4 Deaccentuation condition un-subtracted waves of cl1 cl3 & cl4

4.13.3.1. Results of between clusters comparison of each condition

Results of ANOVA (type 3) between clusters highlight an effect of cluster in the late time-window (570 – 800 ms) in the Deaccentuation condition in the Anterior Left quarter $F(2,20) = 4.84$, $p = 0.003$; in the Anterior Right quarter $F(2,20) = 3.97$, $p = 0.035$; and in the Posterior Right quarter $F(2,20) = 3.71$, $p = 0.043$.
I do not find an interaction effect when considering both the accent conditions and clusters in the late positivity window, and the Deaccentuation condition does not elicit any significant effect in the earlier time-windows. By visual inspection, cl1 shows the tendency to modulate a negativity in the N400 time window and to modulate the positivity in the late-positivity window in the Deaccentuation condition in the frontal electrodes. Moreover, the plots show that cl3 is the subpopulation more sensitive to the modulation of the contrastive focus accent, showing two peaks in the early 180-440 time-window and the most enhanced positivity in the late 570-800 ms time-window.
It is important to highlight that the comparison between clusters is affected by the sample size that after the analysis is one third of the original L2 group, and differences between groups are harder to catch because of the lower statistical power.

4.14. Discussion of the first cluster analysis – external factors and L2 prosody acquisition

The within-cluster analysis considers the modulation of the effects in each cluster, while the between-clusters analysis mainly highly the different modulation of the un-subtracted wave for the Deaccentuation condition and partly of the Contrastive condition across clusters.
Cl1 represents a subpart of the sample with the average lowest AoA and Ao 1st Immersion, a longer formal training, a shorter immersion time and a lower exposure to native spoken language; cl3 reflects an intermediate group with slightly shorter training, compared to cl1, and a slightly higher AoA and Ao 1st Immersion, but who has mainly exposed to native-language speech during the specific training; and cl4 represents a subpart of the sample, with the highest AoA and Ao 1st Immersion, who mainly began to study German later in life because of work-related reasons, who underwent a shorter time of training, but it has been intensively exposed to native-language speech, in the courses and while living in a German-speaking area, since these individuals are the subgroup with the longest immersion time.
Interestingly, comprehensively these 3 types of L2 formations lead to a somewhat equivalent formal level of competences in the certified CEFR levels.
Observing the within-cluster analysis, it is possible to see that the effects are mainly anterior in all 3 groups, but cl4 displays a broader distribution in the modulation of the P200 component associated to the congruent accent condition. The early component in cl4 shows a significant effect also in anterior and posterior regions of the scalp. This main difference in the early component modulation is also highlighted by the fact that cl4 is the only cluster that shows a significant mean difference between the modulation of the congruent condition vs. the other two.
Nonetheless, in the N400 window, the most native-like discrimination between accents in the anterior region is showed by cl1. In cl1, the directions of the effects show to be congruent with the L1 processing and Deaccentuation is the condition displaying the largest negativity. On the contrary, cl4 highlights that the Contrastive Narrow Focus accent generates the highest integration cost. Nonetheless, in none of the separated clusters, the N400 is significant. The congruent vs. non-congruent PA does not emerge with the

smaller sample sizes. The tendency for congruent vs. non-congruent N400 manipulation seems to come mainly from data in cl4, where a tendency is available also with a reduced sample size.
Interestingly the main difference in the effects produced by each cluster emerges in the late-component time-window. Cl1 shows a 3-way modulation of the late-positivity component, higher for the Contrastive accent, followed by Deaccentuation and the congruent manipulation. The effect appears slightly smaller, but significant also in cl3 and becomes 2-way in cl4: contrastive condition vs. congruent and Deaccentuation.
This analysis highlights that cl4 has the strongest auditive response to the prosodic manipulation, in particular, it is the group that responds with the largest extent to H+L* and similarly to the Contrastive Focus accent. In cl1 the early component does not show a significant difference between the congruent and incongruent conditions, but the congruent condition shows the incipit of the process of separation of this modulation as relevantly different from the other two. The component, thus, shows a direction of acquisition that goes from cl4 to cl1 in the display of the modulation of the auditive respond shifting from the pure processing of acoustical salience to the intake considering H+L* as most appropriate accentuation pattern in this experimental setting.
In none of the clusters, the intake is sufficient to trigger a differential integration process linking the accents to the pragmatic structure, indicating that they are equivalently evaluated when the integration with previously processed information takes place.
Strikingly, though, the intake modulation affects the late-component elicitation. The functional interpretation of this component, that has been previously provided, considers the sum of 3 aspects: the response to auditory salience of the peak on the stressed syllable of the final noun in the contrastive condition, the possible decoding of a narrow focus prosodic manipulation, the final discourse's update considering the accentuation pattern and its integration with the content and previously processed information finalized to the integration of the final referent in the model. What it is observable is that from cl4 to cl1 the late-component shifts from a 2-way modulation to a significant 3-way modulation. It is not possible to specifically point to one of the 3 aspects that guide the elicitation of the late -component especially in the contrastive focus condition, but it is possible to provide an explanation. Since cl4 shows the highest early auditory response to acoustical salience and slightly indicates a modulation of the N400 with the highest cost, associated with the contrastive focus accent, it can be hypothesized that the late-component elicitation strongly responds to the perception of an acoustical salient F0 peak on the final stressed syllable, and possibly to the recognition of a contrastive focus accent. Cl4 ultimately indicate that Deaccentuation is never considered a relevant accentuation pattern at any stage of processing.
I suggest therefore that cl4 is the cluster most strongly guided by the accentuation pattern set by the Italian L1 prosodic system.
Cl1, on the other hand, does not show to process the 3 pitch contours completely independently from the native language's prosodic system, but both the early and N400 time-windows display the possible emerging of the target-like processing, even at the time of the experiment, it has not reached a stable representation in their interlanguage. In particular, the N400 window shows a tendency to the 3-way discrimination and integration of the PAs. In the late time window cl1 displays to consider the Deaccentuation pattern as relevant for the final update of the discourse's representation, since it is not possible again here to specifically identify what causes the main positive peak for the contrastive condition, but it is clear that the Deaccentuation pattern requires more cognitive effort than the congruent condition.
This first analysis then highlights that the AoA and Ao 1st immersion primarily guide the construction of L2 target-like processing in L2 learners, supporting the classical position in SLA of "the earlier, the better." In second instance, it positively supports the effect of a longer education that began again earlier in life, even

if not completely dispensed by native-language instructors. There is possibly also a non-directly measured factor playing a role, which is the function of personal motivation in the acquisition of an L2. Cl1 having begun to acquire the L2 earlier in life, and mainly representing people pursuing a high education specifically centered on the acquisition of L2 German, can reflect a group of highly motivated individuals who find a personal satisfaction in the correct mastering of the linguistic system. On the contrary, individuals acquiring an L2 later in life, mainly because of the physical moving in a foreign country for work reasons, can be characterized by a lower motivation specifically directed to the acquisition of specific L2 features and have a higher motivation towards the acquisition of the basic morpho-syntactic structures and vocabulary, that allow a sufficient interaction with native speakers. If Deaccentuation is a marked prosodic structure for L2 Italian learners, in addition, characterized by low acoustical salience, it is possible that its modulation in function of the degrees of Givenness associated to the information status of a referent cannot be properly addressed as relevant to grant a sufficient level of communication, and therefore is not processed as intake. In final instance, this analysis highlights that the most relevant properties guiding the processing of late learners are the acoustical saliency of the peak; therefore, they seem to allocate selective attention in function of prosodic prominence. This supports the view that in Italian the focus manipulation plays a major role in the formation of the discourse structure in the oral communication between interlocutors. Information structure shows to be mainly following the relational the repartition of discourse in focus vs. background information.

4.15. Results of MFA of internal factors (AQ and IRI variables)

The threshold level for each dimension to explain a relevant proportion of variance is 9.09 % therefore, I consider the first 4 dimensions that explain a comprehensive proportion of variance of 74 % (see Table 4.15.1 and Figure 4.15.1)

	Eigenvalue	Explained variance	Cumulative variance
Dim 1	1.98	29.04	29.04
Dim 2	1.28	18.73	47.77
Dim 3	1.05	15.36	63.13
Dim 4	0.74	10.84	73.97

Table 4.15.1 Proportion of exp. Var. in the first 4 Dimensions (MFA analysis AQ-IRI)

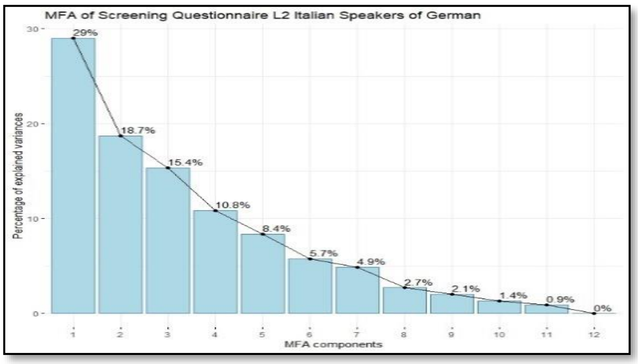


Figure 4.15.1Proportion of explained variance in the first 4 Dimensions (MFA analysis AQ-IRI)

Table 4.15.2 reports the contribution of each group to the dimension.

Groups	Dim1	Groups	Dim2	Groups	Dim3	Groups	Dim4
Gender	28.77	IRI	43.17	AQ	49.70	IRI	46.08
IRI	28.69	AQ	31.75	A	38.22	AQ	29.51
AQ	27.79	A	16.33	IRI	6.41	Gender	16.50
A	14.74	Gender	8.75	Gender	5.67	A	7.91

Table 4.15.2 Group contribution in each Dimension (MFA analysis AQ-IRI)

Table 4.15.3 reports how well each group is represented in the dimension (squared cosine).

Groups	Dim1	Groups	Dim2	Groups	Dim3	Groups	Dim4
Gender	0.33	IRI	0.21	A	0.16	IRI	0.08
IRI	0.22	AQ	0.09	AQ	0.15	AQ	0.03
AQ	0.17	A	0.04	Gender	0.00	Gender	0.00
A	0.09	Gender	0.01	IRI	0.00	A	0.00

Table 4.15.3 Squared - cosine of each Group in each Dimension (MFA analysis AQ-IRI)

The biplot in Figure 4.15.2 reports the variables' cosine values selecting Dim1 and Dim2.

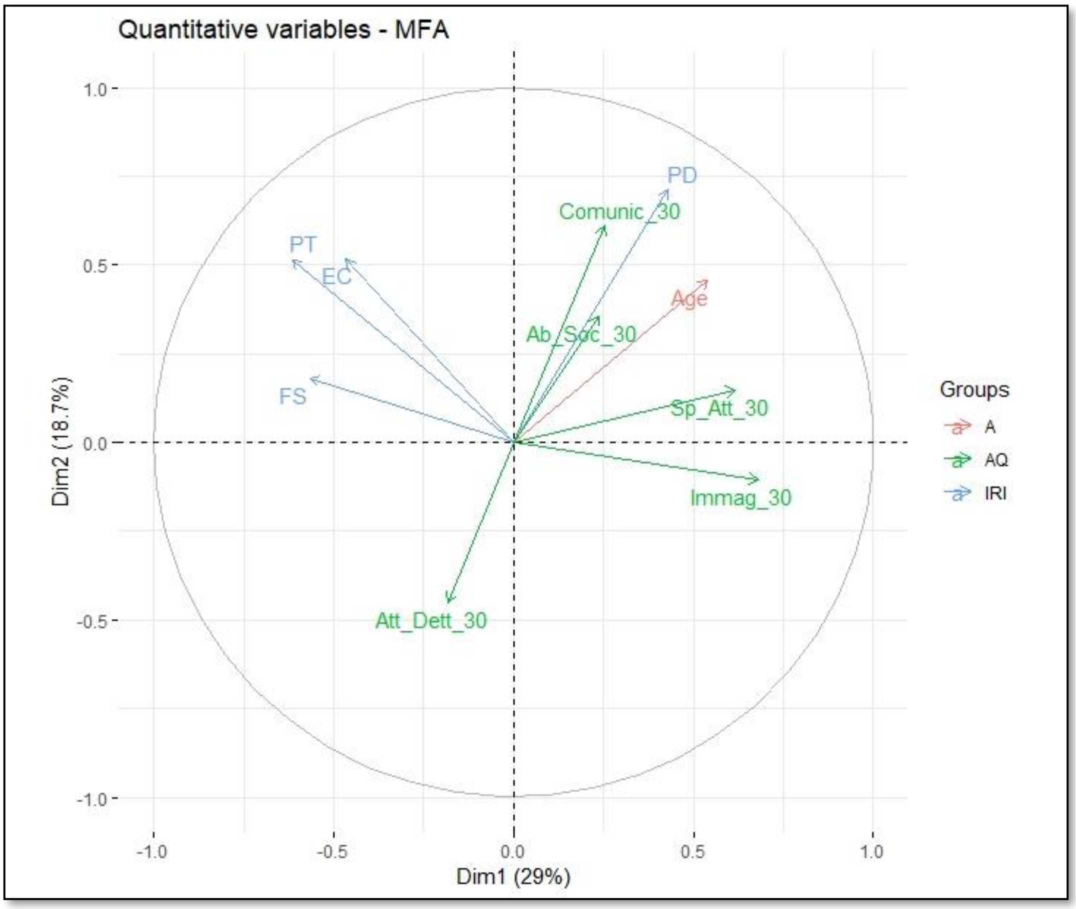


Figure 4.15.2 Biplot of Squared-Cosine of AQ and IRI's variables in Dim1 and Dim2

From the plot above, it is possible to observe that the distribution of the single variables follows the predicted trend. In particular, along Dim1 negative values coherently map together with the positive predisposition of participants to considerate the interlocutors' perspective (PT) and emotional state (EC), and as well their ability to fantasize (FS). The variables measuring their difficulties in socializing (Ab_Soc_30) and communicating (Comunic_30) coherently point to the opposite direction and are also captured by Dim2, resulting therefore in opposition to both the empathic variables (PT & EC) and the variable Att_Dett_30, which measure atypical attention to details of the participants. The variable Immag_30 describing the difficulty of the participants in fantasizing and playing fictional activities is well captured by Dim1 and points to the opposite direction of the variable FS. The variable describing the resistance to the switch of attention (Sp_Att_30) is mapped together with Immag_30 and negatively correlates with the predisposition to fantasize (FS). The variable Age appears to correlate with the variable measuring social impairments (Ab_Soc_30 & Comunic_30), and negatively correlates with the high attention to details (Att_Dett_30).

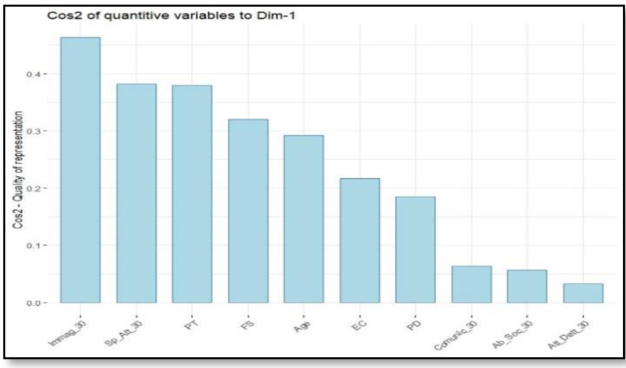


Figure 4.15.4 Squared cosine of quantitative variables in Dim1

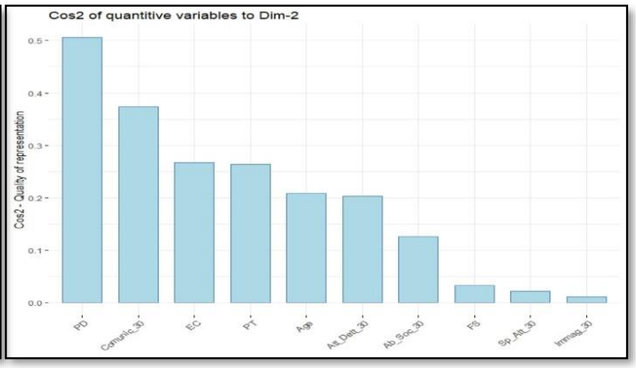


Figure 4.15.3 Squared cosine of quantitative variables in Dim2

I look now on how the MFA organizes the single individuals belonging to the sample and report here single plots relevant both in Dim1 and Dim2 (Figures from 4.15.6 to 415.10)

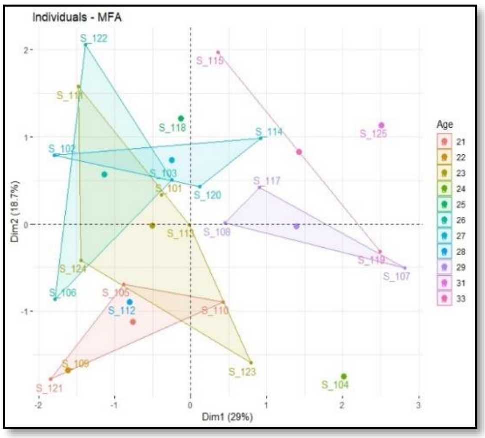


Figure 4.15.5 Age

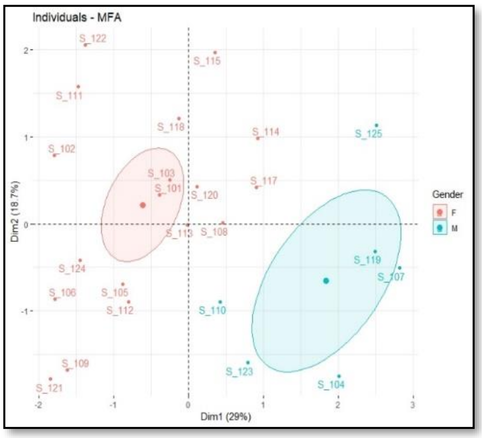


Figure 4.15.6 Gender

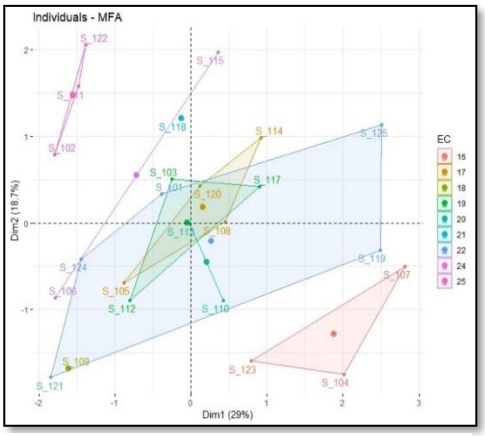


Figure 4.15.8 EC

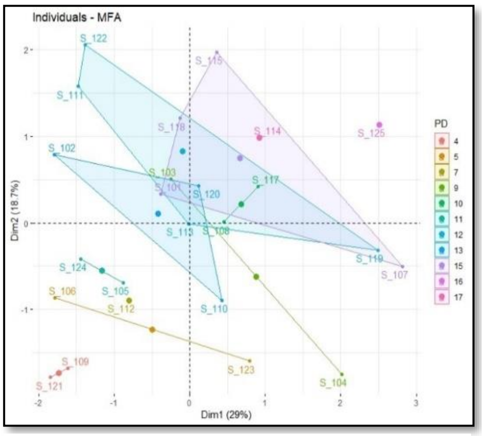


Figure 4.15.7 PD

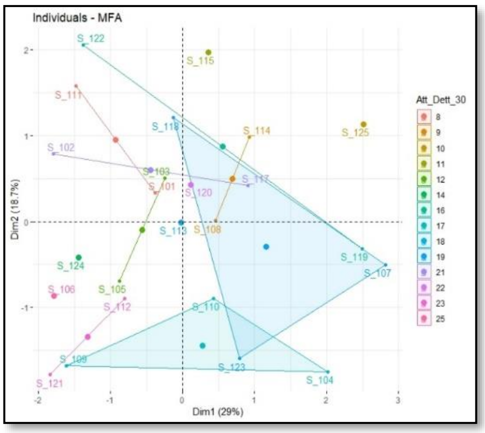


Figure 4.15.10 Att_Dett_30

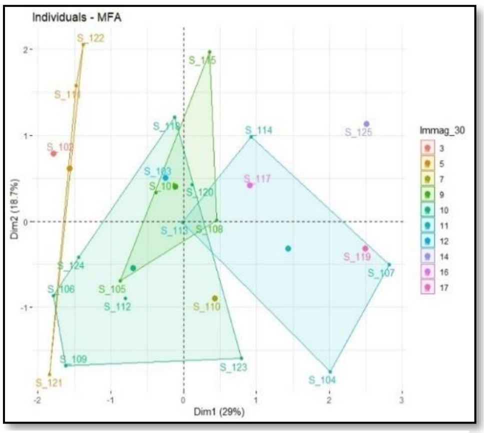


Figure 4.15.9 Immag_30

4.15.1. Cluster Formation

Results of the MFA analysis are taken as input for Hierarchical Clustering on Principal Components (HCPC). The aim is, as in the previous analysis, to observe whether different clusters of the sample, in this case on the basis of AQ and IRI variables can account for processing differences during the perception of specific pitch contours in German L2. In particular, I am interested in observing whether the presence of polarized variables expressing ease in social interactions or on the contrary expressing the presence of distress and of traits in line with the one considered belonging to the autism spectrum correlates with the processing of pitch modulations that are not appropriately matching the underlying pragmatic structure, specifically constructed with bridging inferences. In particular, as in the previous analysis I am interested in the ERP components associated with the processing of the Deaccentuation condition since it is the pitch modulation that more consistently differ between the intonational system in Italian and German.
The results of the HCPC provide the following clusters' subdivision.

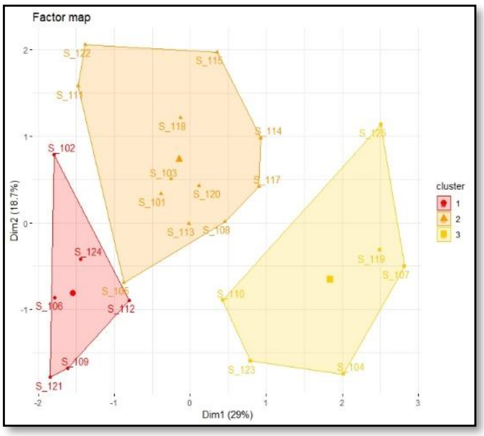


Figure 4.15.11 Cluster formation from MFA analysis of AQ & IRI variables

In the following Table 4.15.4, I report for each cluster, its composition and the most represented dimensions and mean and s.d. of each variable of AQ and IRI. AQ variables are reported without the indication "30" that signals the ceiling value with the adopted scoring method, only for layout reasons.

			mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.
Cl.	**Ind.**	**Dim.**	**FS**	**FS**	**EC**	**EC**	**PT**	**PT**	**PD**	**PD**		
cl1	S_102 S_106 S_109 S_112 S_121 S_124	Dim2 Dim1	20.33	2.25	21.67	2.73	20.00	2.19	7.33	3.83		
			Ab_ Soc	**Ab_ Soc**	**Com.**	**Com.**	**Sp_ Att**	**Sp_ Att**	**Att_ Dett**	**Att_ Dett**	**Imag**	**Imag**
			4.67	1.37	5.17	1.94	9.93	3.31	20.50	4.18	8.00	3.16
cl2	S_101 S_103 S_105 S_108 S_111 S_113 S_114 S_115 S_117 S_118 S_120 S_122	Dim2 Dim4	**FS**	**FS**	**EC**	**EC**	**PT**	**PT**	**PD**	**PD**		
			17.50	3.90	20.25	3.14	20.25	2.96	12.58	2.47		
			Ab_ Soc	**Ab_ Soc**	**Com.**	**Com.**	**Sp_ Att**	**Sp_ Att**	**Att_ Dett**	**Att_ Dett**	**Imag**	**Imag**
			8.50	3.21	11.42	3.12	14.00	2.95	13.75	5.19	9.67	2.93
cl3	S_104 S_107 S_110 S_119 S_123 S_125	Dim1	**FS**	**FS**	**EC**	**EC**	**PT**	**PT**	**PD**	**PD**		
			16.50	2.95	18.67	3.01	16.17	3.19	11.67	4.08		
			Ab_ Soc	**Ab_ Soc**	**Com.**	**Com.**	**Sp_ Att**	**Sp_ Att**	**Att_ Dett**	**Att_ Dett**	**Imag**	**Imag**
			8.50	2.43	9.17	2.64	17.00	3.22	16.33	3.33	11.67	3.44

Table 4.15.4 Clusters composition on the basis of the MFA analysis of AQ & IRI variables

From the table above, it is possible to observe that the FS and EC variables decrease from cl1 to cl3 and that the variable PT has an equal mean value in cl1 and cl2 and lower in cl3. On the other hand, the PD variable is lower in cl1 and then higher but similar in cl2 and cl3. In addition, it is possible to observe that cl1 has mean values below 10 in all AQ variables, but rather high in the variable Att_Dett_30. Cl3, on the other hand, has averagely higher values in all AQ variables, the highest scores within-cluster are the two variables Sp_Att_30 and Att_Dett_30, while cl2 shows to have values in between cl1 and cl3, with no particular emerging scale.

It is important to highlight that the individuals cl1 and cl2 are all females, while all the male population in the sample is in cl3.

These results are in line with the literature observing a main difference in the population between males and females showing that the variables measuring empathy are averagely higher in the female population and the variables belonging to the AQ scales are generally higher in the male population. Two variables appear to be rather gender-independent in this analysis: Att_Dett_30 and PD. Att_Dett_30 mainly discriminates between cl1 and cl2 and shows to be in the same order of dimension between cl2 and cl3, so across gender.

4.16. *ERP - MFA of internal factors (AQ and IRI variables)*

I plot and analyze in this section the modulation of the 3 conditions in each cluster based only on the results of the MFA of AQ and IRI variables with Gender and Age information. I perform here an analogous analysis of the clusters' differences looking at the main effects in early medial and late time-window. I am interested in observing whether the presence of individual variability in the processing of pragmatic and social information during oral communication affects the acquisition of the prosodic marking in L2. In particular, whether higher values of traits belonging to the characteristics of the Autism Spectrum affects the acquisition of pitch modulation in function of both hierarchical organization of Givenness and presence of narrow focused elements. Given the results of the previous group analysis, it is of particular interest the modulation of the variables pertaining attention in the AQ test: Att_Dett_30, indicating a usually higher than the neurotypical allocation of attention towards details and particulars within the perceived complex stimulus and Sp_Att_30, indicating the resistance in attention switching during engagements in activities.
Together with the previously presented cluster analysis highlighting relevant acquisitional features I want to delineate a possible profile of a learner acquiring with rather ease compared to the average population's trend and a possible profile of a learner displaying a higher degree of difficulty and impairment in the L2 acquisition of prosodic marking. As in the previously presented cluster analysis, I report plots and analysis for each separate cluster plotted with the same color's association and dashed and dotted lines, and in second instance, the EEG waveform of the un-subtracted waves in each condition between clusters are considered.

4.16.1. Analysis of effects within each cluster

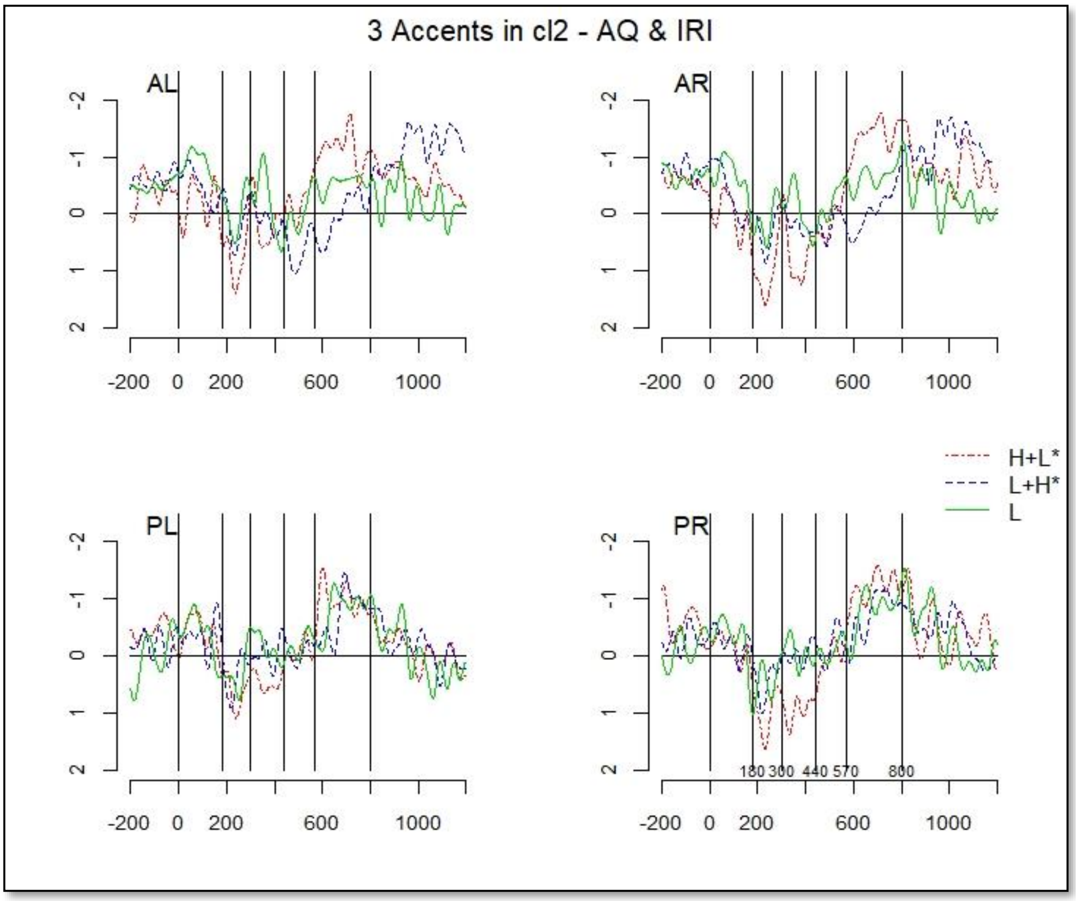


Figure 4.16.1 4 ROI - Cluster2 (MFA analysis of AQ & IRI's variables)

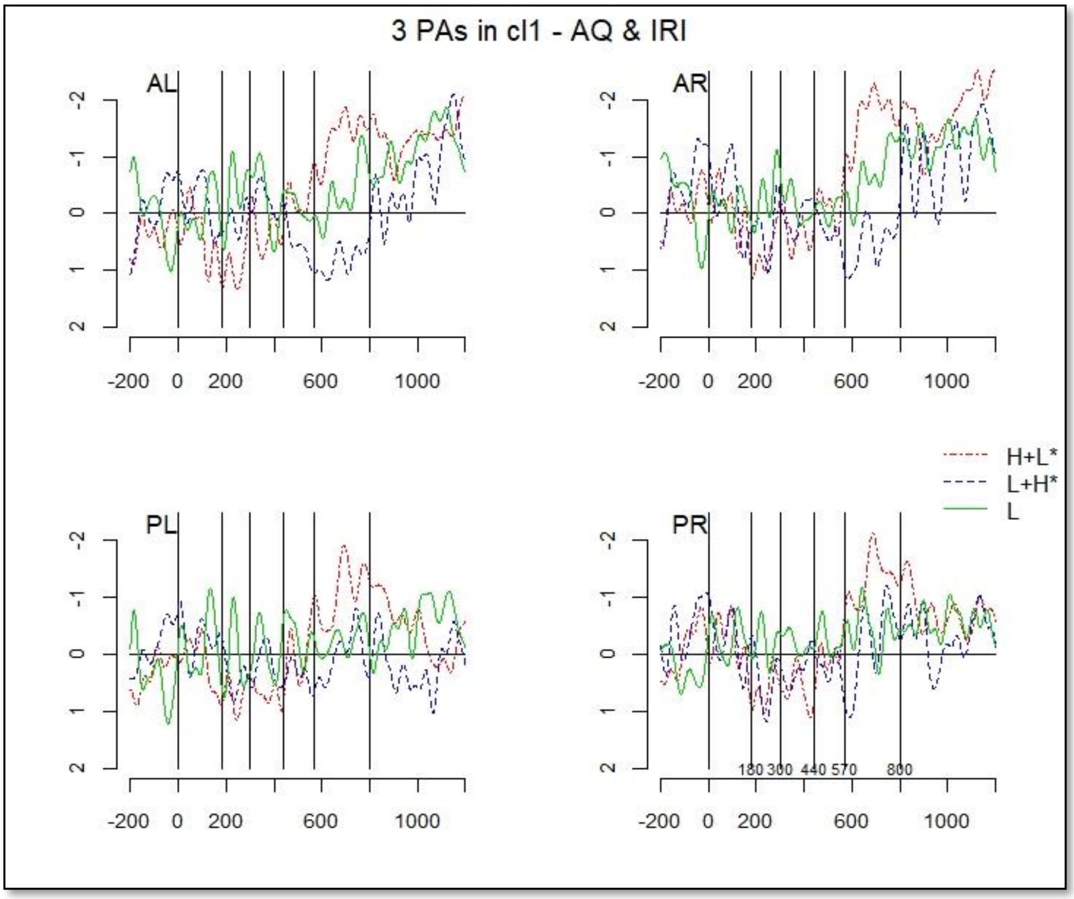


Figure 4.16.2 Cluster1 (MFA analysis of AQ & IRI's variables)

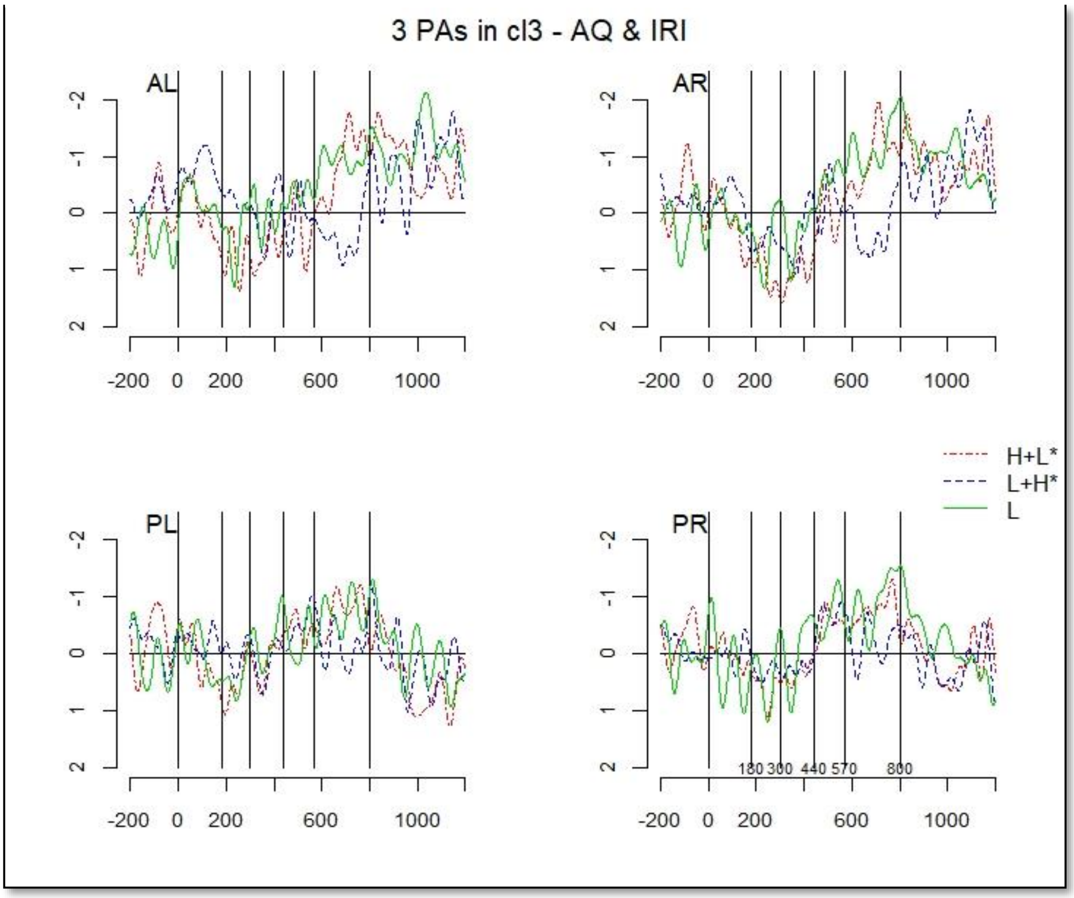


Figure 4.16.3 Cluster3 (MFA analysis of AQ & IRI's variables)

4.16.2. Results of main factor PROSODY and lateral topographic distribution

4.16.2.1. Cl1 - Lateral

Over lateral electrodes, in the 180 – 300 ms time window, I find an effect for PROSODY $F(2,10) = 10.74$; $p = 0.003$. In the 300 – 440 ms time window, the data show only a tendency for PROSODY $F(2,10) = 3.38$; $p = 0.08$; and in the 570 – 800 ms time window, both an effect for Prosody $F(2,10) = 28.87$; $p = 0.0001$; and an interaction of the topographic factor with the main prosody effect $F(6,30) = 6.17$; $p = 0.0003$ are present.

4.16.2.2. Cl2 - Lateral

Over lateral electrodes, in the 180 – 300 ms time window, the data show a main effect for PROSODY $F(2,22) = 4.46$; $p = 0.02$. In the 300 – 440 ms time window, I find an effect for PROSODY $F(2,22) = 3.71$; $p = 0.04$; and a tendency for the interaction between the topographic factor and the main prosody effect $F(6,66) = 1.98$; $p= 0.08$. In the 570 – 800 ms time window, I find both an effect for Prosody $F(2,22) = 7.72$; $p = 0.003$; in addition, the data show a tendency of a topographic effect of the Lateral factor $F(3, 33) = 2.39$; $p = 0.09$; and an interaction with the main prosody effect $F(6,66) = 3.36$; $p = 0.006$.

4.16.2.3. Cl3- Lateral

Over lateral electrodes, in the 180 – 300 ms time window, the data show only a tendency of the topographic factor $F(3,15) =$; $p = 0.09$. In the 300 – 440 ms time window, the results highlight only an effect of the topographic factor $F(3,15) = 4.64$; $p =0.02$. In the 570 – 800 ms time window, I find both an effect for Prosody $F(2,10) = 10.38$; $p = 0.004$; in addition, the data show an interaction of the topographic factor with the main prosody effect $F(6,30) = 5.00$; $p = 0.001$.

Pairwise Comparisons in 4 ROI for each cluster

In cl1, pairwise comparisons FDR corrected for 3 within-subjects contrasts, highlight an anterior effect in the 180-300 ms time window, and a broadly distributed effect, significant in all 4 ROI in the late-positivity window.

In Anterior Left quarter, between 180-300 ms, the data show a tendency in two comparisons: Congruent vs. Deaccentuation $t(5) = 3.03$; $p = 0.05$; C.I.(0.19 – 2.31); Contrastive and Deaccentuation $t(5) = -2.85$; $p = 0.05$; C.I (-0.87 - -0.04). In Anterior Right quarter, between 180-300 ms, I find a tendency in the same two comparisons: Congruent vs. Deaccentuation $t(5) = 2.85$; $p = 0.08$; C.I.(0.1 – 1.89); Contrastive and Deaccentuation $t(5) = -2.50$; $p = 0.08$; C.I (-1.17 - 0.02). In the 570 – 800 ms time window the factor PROSODY shows a significant mean difference:

In the Anterior Left quarter: there is a significant effect of PROSODY in all 3 comparisons: Congruent and Contrastive accent $t(5) = -5.90$; $p = 0.006$; C.I.(-1.45 - -0.57); Congruent vs. Deaccentuation condition $t(5) = -4.45$; $p = 0.01$; C.I.(-3.46 - -9.92); and Contrastive vs. Deaccentuation $t(5) = -2.49$; $p = 0.05$; C.I.(-2.40 - -0.04).

In the Anterior Right quarter: there is a significant effect of PROSODY in all 3 comparisons: Congruent and Contrastive accent t (5) =-4.20; p = 0.01; C.I.(-1.64 - -0.39); Congruent vs. Deaccentuation condition t (5) = -4.09; p = 0.01 ; C.I.(-3.57 - -0.84); and Contrastive vs. Deaccentuation t (5) = -3.43; p = 0.02; C.I.(-2.16 – 0.31).
In the Posterior Left quarter: there is a significant effect of PROSODY in 2 comparisons: Congruent and Deaccentuation accent t (5) = -3.60; p = 0.02; C.I.(-1.46 - -0.25); Congruent vs. Contrastive condition t (5) = -3.86; p = 0.02; C.I.(-1.86 - -0.37);
In the Posterior Right quarter: there is a significant effect of PROSODY in the same 2 comparisons: Congruent and Deaccentuation accent t (5) = -4.50; p = 0.02; C.I.(-1.40 - -0.38); Congruent vs. Contrastive condition t (5) = -3.76; p = 0.02; C.I.(-1.96 - -0.37).

In cl2, pairwise comparisons FDR corrected for 3 within-subjects contrasts, highlight an anterior right effect in the early time window, a right lateralized effect in the 300-440 ms time window, and an anterior effect in the late-positivity time window.
In Anterior Right quarter, between 180-300 ms, there is a significant effect of PROSODY in one comparison: Congruent vs. Deaccentuation t (11) = 4.30; p = 0.004; C.I.(0.46 – 1.43); a tendency in the: Congruent vs. Contrastive t (11) = 2.19; p = 0.08; C.I.(-0.006 – 1.53); and no significant mean difference between the Contrastive and Deaccentuation condition.
In Anterior Right quarter, between 300-440 ms, a significant effect of PROSODY is found only in one comparison: Congruent vs. Deaccentuation t (11) = 3.03; p = 0.03; C.I.(0.21 – 1.31);
In Posterior Right quarter, between 300-440 ms, the data show a significant effect of PROSODY in two comparisons: Congruent vs. Deaccentuation t (11) = 2.72; p = 0.05; C.I.(0.19 – 1.80); Congruent vs. Contrastive t (11) = 2.44; p = 0.05; C.I.(0.1 – 1.86); and no significant mean difference between the Contrastive and Deaccentuation condition.
In the 570 – 800 ms time window the factor PROSODY shows:
In the Anterior Left quarter: a tendency for the factor PROSODY in one comparison: Congruent vs. Deaccentuation t (11) = -2.14; p = 0.09; C.I.(-1.20 – -0.02); and effect in the Congruent vs. Contrastive accent t(11) = -3.08; p = 0.03; C.I.(-2.00 - -0.33);
In the Anterior Right quarter: an effect for the factor PROSODY in one comparison: Congruent vs. Deaccentuation t (11) = 4.30; p = 0.004; C.I.(0.46 – 1.43); and a tendency in the Congruent vs. Contrastive accent t (11) = 2.19; p = 0.08; C.I.(-0.005 – 1.53).

In cl3, pairwise comparisons, FDR corrected for 3 within-subjects contrasts, highlight only an anterior effect in the 570-800 ms time window.
In the 570 – 800 ms time window the factor PROSODY shows:
In the Anterior Left quarter: an effect in the Congruent vs. Contrastive comparison t (5) = -3.55; p = 0.05; C.I.(-2.12 - -0.34); and in Contrastive vs. Deaccentuation comparison t (5) = -2.93; p = 0.05; C.I.(-2.44 - -0.16).
In the Anterior Right quarter: an effect in the same two comparisons: Congruent vs. Contrastive comparison t (5) = -6.40; p = 0.004; C.I.(-1.75 - -0.75); and in Contrastive vs. Deaccentuation comparison t (5) = -3.97; p = 0.02; C.I.(-2.38 - -0.51).

4.16.3. Between clusters comparison of each condition

Since the greatest difference in the sample is between cl1 and cl3, the EEG waveform of these two subparts is plotted in Figures 4.16.4, 4.16.5 and 4.16.6. EEG waveforms are reported for each accent in association with cl1 (H+L* - red, Contrastive – blue, Deaccentuation – green) and cl3 is always plotted in purple and with a dashed line. Nonetheless, cl2 is introduced in the analysis below in order not to bias the results.

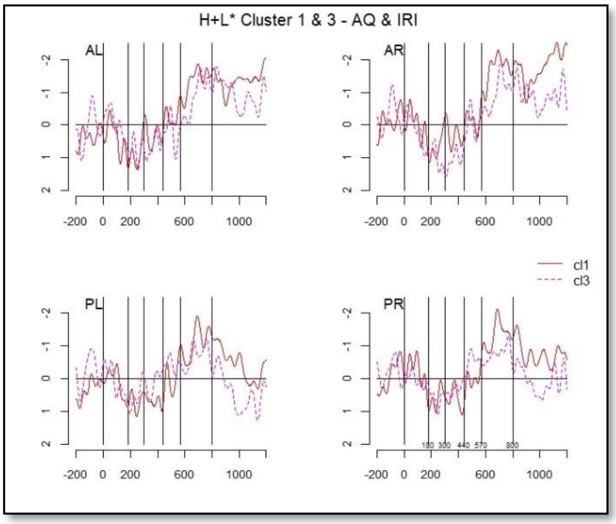


Figure 4.16.4 Congruent condition's AQ & IRI's Cluster1 and Cluster3

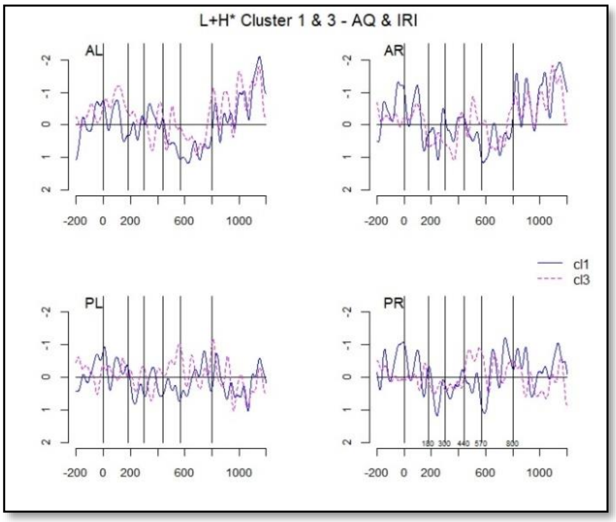


Figure 4.16.5 Contrastive condition's AQ & IRI's Cluster1 and Cluster3

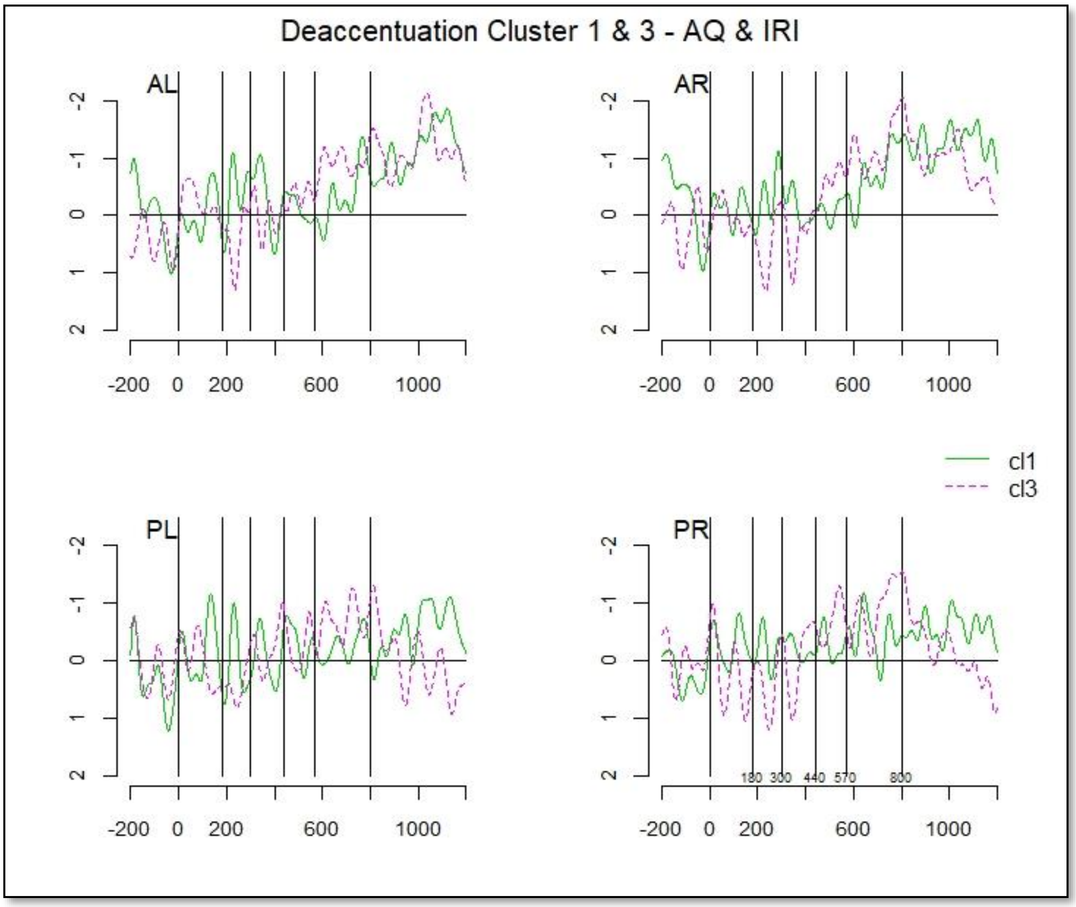


Figure 4.16.6 Deaccentuation condition's AQ & IRI's Cluster1 and Cluster3

The results of ANOVA (type 3) comparison highlight only a tendency, for the factor CLUSTER in the Deaccentuation condition in the Anterior Right quarter between 180 and 300 ms $F(1, 21) = 1.83$, $p < 0.2$.

4.17. Discussion of the cluster-analysis considering AQ and IRI variables

The presented cluster analysis based on Gender, Age and AQ and IRI variables produced 3 clusters characterized by a balancing effect of AQ and IRI variables. The literature on the distribution and validation of the IRI test and its variable highlights the possibility that the PD variable is of a significantly different nature of the other 3 (FS, EC, PT) and the data seems to confirm across clusters this data explanation. In fact, the 3 remaining variables of the IRI measure 3 dimensions that signal the presence of the empathic traits in the individual, while the PD variable signals a sort of condition of anxiety and distress of the participant in critical situations that intuitively recalls the presence of a difficulty in the participant rather than a facilitatory effect in the handling of the social situations. If we look at mean values per cluster of the IRI dimensions indicating the presence of the empathy construct in the individual (FS, EC, PT) we see that the mean values decrease shifting from cl1 to cl3: mean value of FS, EC, & PT in cl1= 20.7; mean value of FS, EC, & PT in cl2 = 19.3; mean value of FS, EC, & PT in cl3 = 17.1. Looking at the AQ variables we see that Att_Dett_30 has a majoring discriminating function between clusters and appears to modulate differently from the other AQ variables, therefore, excluding it from an averaging process we see that mean values of the remaining 4 AQ dimensions (Ab_Soc_30; Communic_30; Sp_Att_30; Imag_30) show an increasing trend from cl1 to cl3: mean values in cl1 = 6.9; mean value in cl2 = 10.9; mean value in cl3 = 11.6.
Thus, cl1 is characterized by a stronger presence of “empathic concern” and a lower presence of traits belonging to ASC, and on the contrary, cl3 is characterized by a stronger presence of ASC traits and a lower presence of “empathic concern”’ traits. As previously said the clusters are polarized in gender differences since all the male population in the sample has been position in cl3. Att_Dett_30, measuring the individual characteristic of paying overall higher attention than the average population to the details in a visual and auditive scene, and to dates and the number shows to modulate across clusters orthogonally compared to Empathic and the remaining AQ variables. In fact, the mean value in the clusters is not organized along the same trend: cl1 = 20.50, cl2 = 13.75, cl3 = 16.33. Thus, when comparing cl2 and cl3, Att_Dett_30 appears to modulate coherently with the other AQ variables, but when comparing cl1 with cl3, the results are quite anomalous.
If we look at ERP elicited components across clusters, we see that: cl1 and cl2 display a comparable modulation of the early P200 component, where mainly the congruent and contrastive conditions are disentangled from the Deaccentuation condition. Cl2 compared to cl1 also shows a significant effect in the N400 modulation with a right-lateralized effect, that distinguishes the congruent condition from the other 2. Both cl1 and cl2 show a 3 -way modulation of the late-component but cl1 display a broad distribution over the scalp while cl2 only a main frontal effect.
Cl4’s ERP modulation clearly shows no early P200, nor the intermediate N400 effect, and only a two-way modulation of the late-positivity responding only to the acoustically salient L+H* peak.
From this analysis, it is plausible to conclude that the most advanced group in the processing of L2 prosodic marking in function of the information structure of discourse is cl2.
Cl2 shows the emerging of a 3-way modulation of the effect both in the N400 and late-positivity component, with a strong right lateralization in the two early and intermediate component and also the left anterior

modulation of the late-positivity. Cl2 is a subgroup of individuals having the central values of the distribution of both the scores in AQ and IRI relevant variables; it appears therefore that the advantaging and disadvantaging properties of both the empathic traits and ASC traits are equally compensate.
On the contrary cl1 shows a deeper modulation of the late-positivity compared to cl2, no modulation of the N400 component, and tendency in the identification of the congruent prosodic pattern distinguished from the other 2. Cl1 has a stronger presence of empathic traits, the lowest values in AQ variable, and the highest values in Att_Dett_30. This highlights that this group of individuals reports to strongly rely on the perspective-taking mechanism in social situations, nonetheless, the N400 modulation is not significant. What happens in the other hand, is that they appear to be the cluster that mostly largely responds to the auditive perception of the 3 prosodic modulations. In particular, in the left anterior quarter, the P200 modulation shows the trends towards a 3-way auditive response. This is strikingly in line with the functional interpretation of the P200 as an attentional component, and with the language-aptitude concept of intake. Very interestingly the auditory discrimination is probably not yet linked to the manipulation of Givenness in German L2; therefore Deaccentuation is not considered a relevant cue in the integration of the discourse structure, but this cluster then evidently show to consider a 3-way modulation of pitch distributed all over the scalp. The broad distribution could suggest that this subgroup of individual computes to a stronger extent compared to the other individuals in the L2 group, a processing of the discourse update associating the prosodic contour to a degree of informativeness of the final referent more closely related to the German L1 effect than to the response of the acoustical salience of the L+H* accent.
Cl3, on the other hand, shows a clear disadvantage in the processing of the 3 PAs and what the data suggest is that this group of individuals only acoustically responds to the prominent peak aligned to the final stressed syllable in the contrastive focus condition.
Summarizing this analysis, it is possible to say that the best individual profile for the processing of L2 prosodic contours associated with the underlying discourse structure is the compensatory effect of the presence of individual traits highlighting a strong presence of empathic and ToM traits and traits belonging to the AQ. This appears to create a condition of equilibrium that leads to the assumption that the external factors pertaining to the period of training and immersion in L2 speaking area have a higher impact. On the other hand, only the strong presence alone of Empathy and Perspective Taking dimensions does not indicate a more refined and connected association between intonation and discourse structure in L2; otherwise, we should have seen a much larger N400 3-way modulation in cl1 compared to cl2. Nonetheless, the presence of overtly high attention allocation to acoustic stimuli can enhance and stimulate the intake processing, and this impacts in later stages the quality of the “wrapping-up” computation of the whole discourse. Cl1 is the subgroup of the population that mostly considers all aspects in the late-component window. Finally, the only presence of ASC traits clearly shows to be a disadvantage for both the computation of the shift from input to intake and for the association of prosody to pragmatics. This is also reflected in the late-positivity window where it is presumable that in the presence of the original not strikingly salient H* aligned to the final syllable any effect would have been absent in all 3 time-windows. Within this respect, it is not possible not to notice that these results indicate a gender advantage of female participants compared to male participants in the acquisition L2 of prosodic properties, extendable probably to the processing of pragmatic aspects of language.

4.18. *Final Cluster Formation: Easiness vs. Impairments in the L2 acquisition of prosodic marking*

In the two previously presented analyses, we have seen in each cluster formation that the following variables play a determinant role in the acquisition of a 3-way processing modulation of the 3 different accent-conditions. Namely, we have seen in the screening-variables analysis that a determinant role in the clusters separation is played by the Age of Acquisition and Age of 1st immersion, corresponding to the literature concept of Age of Arrival and the length of immersion comprehending the exposure to native-speaker speech. In the second cluster-analysis, we have seen that the relevant variables discriminating within the sample are the combination of high level of Att_Dett_30 with low values in the remaining AQ variables, in particular, low values of the variable SP_Att_30; and in addition, higher values in the Empathic concern's variables of the IRI: EC and PT.
Overall, the analysis suggests that the combination of this properties should let emerge the profile of somewhat more advanced prosody L2 learner, compared to the opposite modulation (higher age of acquisition and of Arrival, high AQ variables and particularly low Att_Dett_30 values, in combination with low empathic concern's variables: EC and PT).
Also, although the sample is not balanced for gender's comparison, it appears that the combination of these variables suggests a female advantage in the acquisition of prosody.
Therefore, I perform a final cluster analysis introducing only a combination of the principal variables emerged from the screening variable analysis and the principal one emerging from the analysis of pragmatic traits. I consider, therefore: Gender, Age of Acquisition, Age of 1st immersion, length of immersion at the time of the experiment and two variables from the AQ test: Att_Dett_30 and Sp_Att_30 and 2 variables from the IRI scale: EC and PT. The results of the cluster analysis are primarily functional to the final ERP comparison. I want thus compare individuals with a combination of traits that overall can promote and represent an advantage for the acquisition of L2 prosody and individuals with overall a combination of traits correlating with a less native-like processing that prevent already in the P200 response to consider the 3 modulations as different intakes for the following stages of the discourse processing. Results of the cluster analysis are therefore aggregate to form only 2 main clusters.

4.18.1. Results of the cluster analysis

The threshold level for each dimension to explain a relevant proportion of variance is 12.02 % therefore, I consider the first 4 dimensions that explain a comprehensive proportion of variance of 69 %. (Table and Figure 4.18.1 below)

	Eigenvalue	Explained variance	Cumulative variance
Dim 1	2.06	24.29	24.29
Dim 2	1.49	17.55	41.85
Dim 3	1.19	14.00	55.85
Dim 4	1.12	13.18	69.03

Table 4.18.1 Proportion of explained variance in final cluster analysis

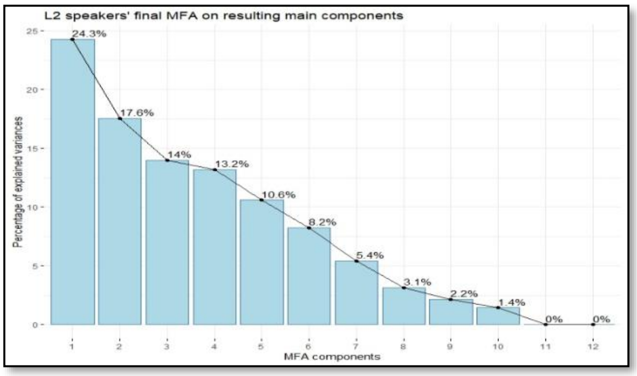


Figure 4.18.1 Proportion of explained variance in final cluster analysis

Table 4.18.2 reports the contribution of each group to the dimension.

Groups	Dim1	Groups	Dim2	Groups	Dim3	Groups	Dim4
Gender	33.24	Immersion	50.52	Immersion	68.39	Immersion	48.78
IRI	23.53	AQ	25.87	AQ	9.59	AQ	28.16
Age	21.24	Age	17.45	IRI	5.25	Age	16.58
AQ	15.07	IRI	5.84	Age	4.88	IRI	5.26
Immersion	6.92	Gender	0.33	Gender	1.89	Gender	1.22

Table 4.18.2 Contribution of each group to each Dimension (Final Cluster analysis)

Table 4.18.3 reports how well each group is represented in the dimension (squared cosine).

Groups	Dim1	Groups	Dim2	Groups	Dim3	Groups	Dim4
Gender	4.7E-01	Immersion	1.9E-01	Immersion	2.2E-01	Immersion	9.9E-02
IRI	2.2E-01	AQ	8.9E-02	Gender	2.0E-02	AQ	6.0E-02
Age	1.7E-01	Age	5.8E-02	AQ	7.8E-03	Age	3.0E-02
AQ	5.8E-02	IRI	7.0E-03	IRI	3.6E-03	IRI	3.2E-03
Immersion	6.8E-03	Gender	2.4E-05	Age	2.9E-03	Gender	1.9E-04

Table 4.18.3 Squared Cosine of each group to each Dimension (Final Cluster analysis)

The biplot in Figure 4.18.2 reports the variables' cosine values selecting Dim1 and Dim2.

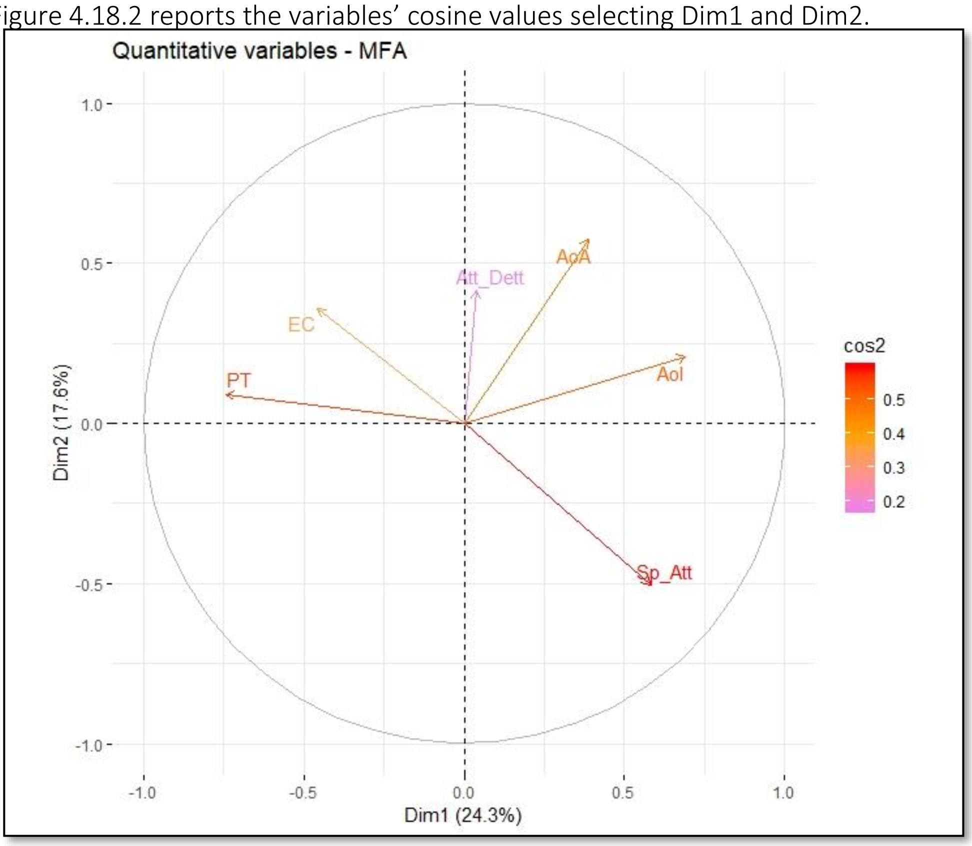


Figure 4.18.2 Biplot of Dim1 and Dim2 (Final Cluster analysis)

Figures from 4.18.5 to 4.18.12 show the sample distribution of single variables relevant both in Dim1 and Dim2.

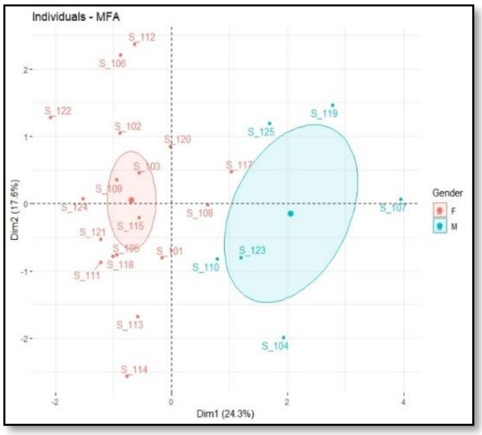


Figure 4.18.3 Gender

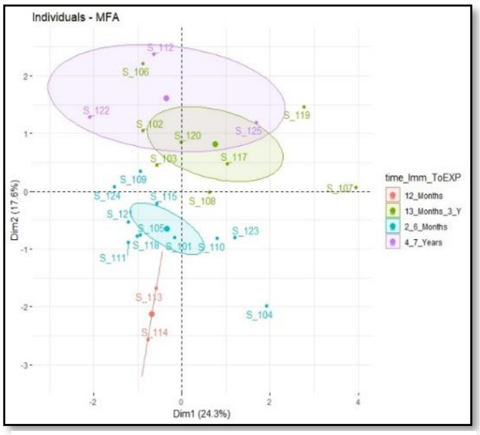


Figure 4.18.4 Length of Immersion

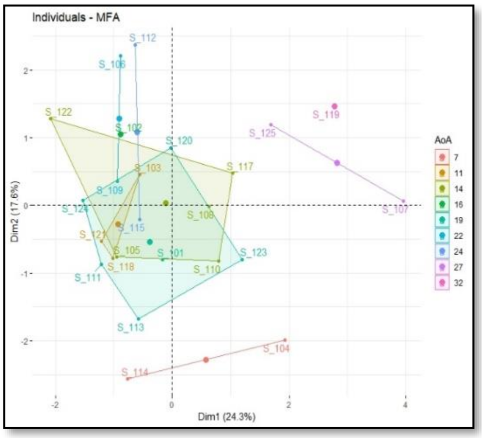


Figure 4.18.5 Age of Acquisition

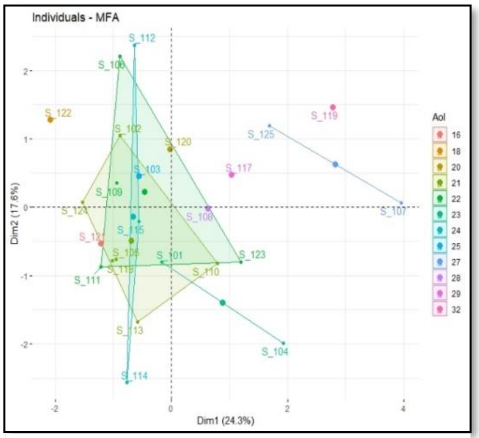


Figure 4.18.6 Age of 1st Immersion

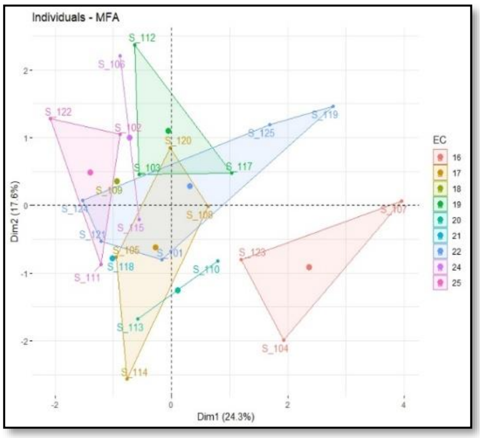


Figure 4.18.8 EC

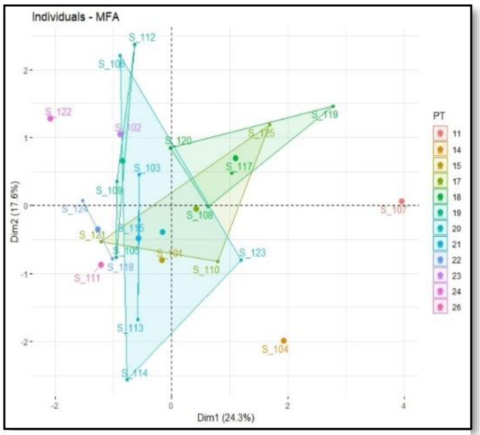


Figure 4.18.7 PT

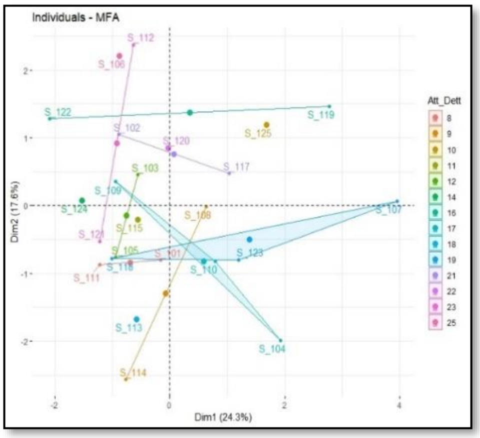


Figure 4.18.10 Att_Dett

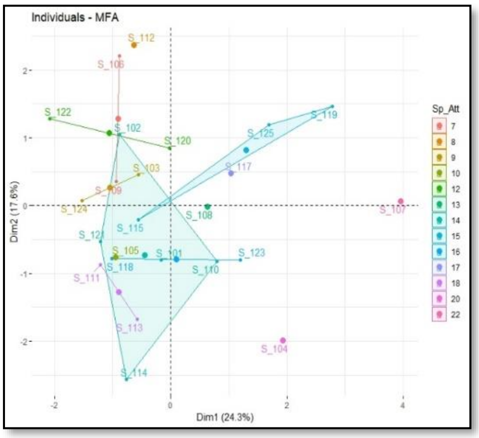


Figure 4.18.9 Sp_Att

The MFA final analysis shows that the two main variables creating a clear cut in the L2 group are Gender and the length of immersion. Individuals positioned in the right bottom quarter of the plots are male individuals displaying the highest values in the variable Sp_Att_30 (poor attention switching) and the lowest in the variables EC and PT. On the contrary, on the top left of the plots are female individuals, declaring the most prolonged period of immersion, the highest values of EC and PT. Overall the disposition of the individuals apart from Gender differences seems not to polarize across these two dimensions but to gradually shift from the top left area of the plots to the bottom right. This is coherent with the previous analysis since we are merging in this final cluster formation the principal components emerging from the previous subdivisions. It is possible that single individual profiles comprehend parts of the factors polarizing the clusters in the first classification along the dimensions evaluating the training in L2, and parts of the factor guided by ASC and empathic traits. In the overall sample distribution, individuals, clearly showing the polarized aspects opposing the two final clusters, must be overall few corresponding to the tails of the distribution.

4.18.2. Cluster Formation

Results of the MFA analysis are taken as input for Hierarchical Clustering on Principal Components (HCPC). The results of the HCPC provide the following clusters' subdivision.

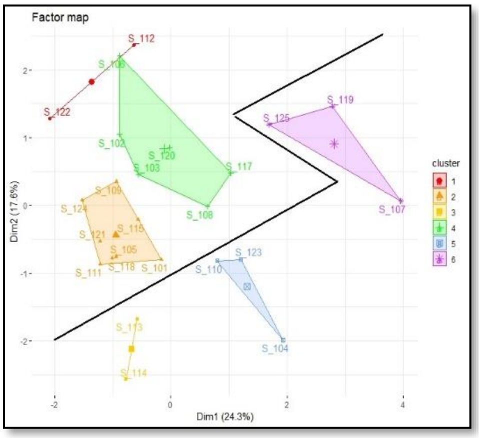


Figure 4.18.11 Final Clusters

As predicted, the cluster formation decomposes the sample in more than two clusters since they show a combination of values that rather gradually shift from the two opposite tails of the distribution. Since the aim is to compare the ERP results, the formation of several smaller groups is not suitable for a final analysis, because it implies that the grand-average procedure aggregating the signal recorded from participants in each cluster would be computed on a small number of participants and the signal would remain too noisy to perform a group comparison. Therefore, the participants are divided aggregating the clusters coherently belonging to the direction of the effects displayed by single variables, trying to compare two groups having the closest possible sizes. The analysis, therefore, indicates that cl1 cl2 and cl4 share a common modulation of all the variables and that they are opposed to cl3 cl5 and cl6. The analysis continues comparing the EEG signal of these two final clusters.

In the following Tables 4.18.4 and 4.18.5 the final placement of each participant in one of the two clusters is displayed, along with values in the relevant variables.

Cluster 1 (cl1)

	cluster	Gender	AoA	Ao 1st Imm.	time_Imm_ToEXP	Att_Dett	Sp_Att	EC	PT
S_101	1	F	19	23	2_6_Months	8	14	22	15
S_102	1	F	16	21	13_Months_3_Y	21	14	25	23
S_103	1	F	11	25	13_Months_3_Y	12	9	19	21
S_105	1	F	14	21	2_6_Months	12	10	17	19
S_106	1	F	22	22	13_Months_3_Y	25	7	24	20
S_108	1	F	14	28	13_Months_3_Y	9	13	17	18
S_109	1	F	22	22	2_6_Months	17	7	18	19
S_111	1	F	19	22	2_6_Months	8	18	25	26
S_112	1	F	24	24	4_7_Years	23	8	19	19
S_115	1	F	24	24	2_6_Months	11	15	24	21
S_117	1	F	14	29	13_Months_3_Y	21	17	19	18
S_118	1	F	11	21	2_6_Months	19	16	21	22
S_120	1	F	19	20	13_Months_3_Y	22	12	17	18
S_121	1	F	11	16	2_6_Months	23	14	22	17
S_122	1	F	14	18	4_7_Years	16	12	25	24
S_124	1	F	19	21	2_6_Months	14	9	22	22

Table 4.18.4 Final Comparison - Cluster1

Cluster 2 (cl2)

	cluster	Gender	AoA	Ao 1st Imm.	time_Imm_ToEXP	Att_Dett	Sp_Att	EC	PT
S_104	2	M	7	23	2_6_Months	17	20	16	14
S_107	2	M	27	27	13_Months_3_Y	19	22	16	11
S_110	2	M	14	21	2_6_Months	17	14	20	17
S_113	2	F	19	21	12_Months	18	18	20	21
S_114	2	F	7	24	12_Months	9	14	17	20
S_119	2	M	32	32	13_Months_3_Y	16	15	22	18
S_123	2	M	19	22	2_6_Months	19	16	16	20
S_125	2	M	27	27	4_7_Years	10	15	22	17

Table 4.18.5 Final Comparison - Cluster2

4.18.3. Cluster's descriptive analysis

Table 4.18.6. and Table 4.18.7 show the values of descriptive statistics of the two final clusters, highlighting the presence of significant differences in the values of relevant variable between the two groups of participants. In particular the two clusters diverge for the values of the variables corresponding to the Empathy construct.

Cluster 1						
Variable	Mean	S.D.	Var.	Min	Max	Obs.
AoA	17.06	4.54	20.60	11	24	16
Ao 1st Imm.	22.31	3.28	10.76	16	29	16
Att_Dett_30	16.31	5.84	34.10	8	25	16
Sp_Att_30	12.19	3.53	12.43	7	18	16
EC	21.00	3.03	9.2	17	25	16
PT	20.13	2.83	7.98	15	26	16

Table 4.18.6 Descriptive Statistics of the values of the most relevant variables in Final Cluster1

Cluster 2						
Variable	Mean	S.D.	Var.	Min	Max	Obs.
AoA	19.00	7	87.14	7	32	8
Ao 1st Imm.	24.63	3.85	15.55	21	32	8
Att_Dett_30	15.63	3.93	15.41	9	19	8
Sp_Att_30	16.75	2.96	8.79	14	22	8
EC	18.63	2.67	7.13	16	22	8
PT	17.25	3.37	11.38	11	21	8

Table 4.18.7 Descriptive Statistics of the values of the most relevant variables in Final Cluster2

In Figures from 4.18.15 to 4.18.19 the clusters' aggregated data of the relevant variables are plotted.

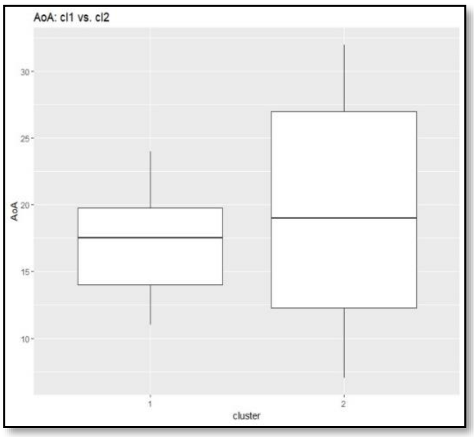


Figure 4.18.13 *Age of Acquisition in Cluster1 and Cluster2*

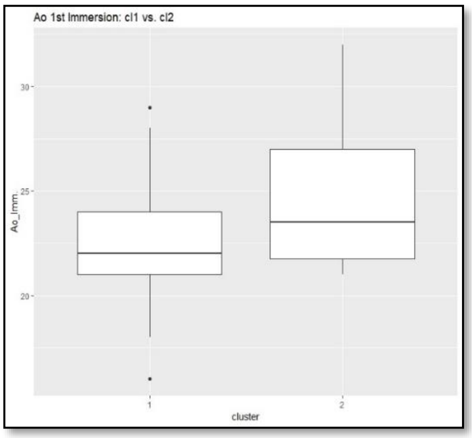


Figure 4.18.12 Age of 1st Immersion in Cluster1 and Cluster2

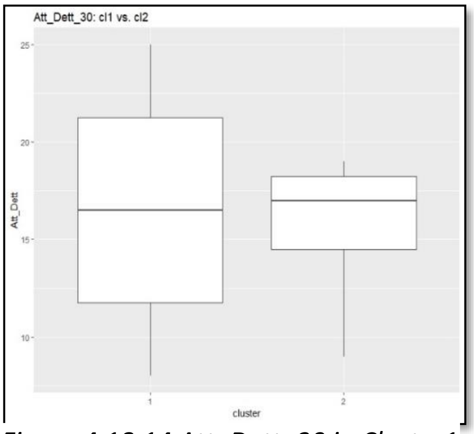


Figure 4.18.14 Att_Dett_30 in Cluster1 and Cluster2

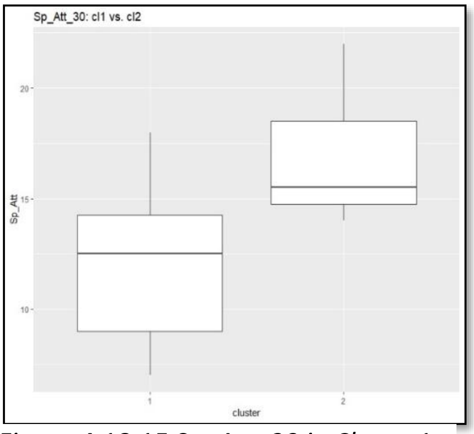


Figure 4.18.15 Sp_Att_30 in Cluster1 and Cluster2

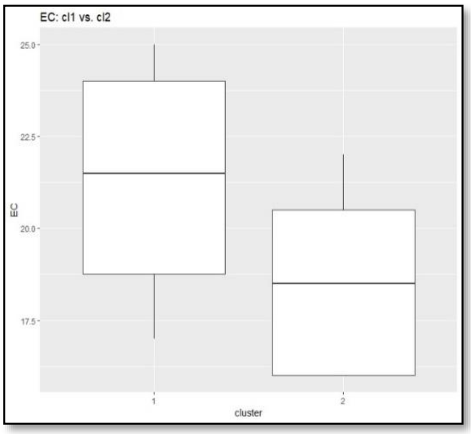


Figure 4.18.17 EC in Cluster1 and Cluster2

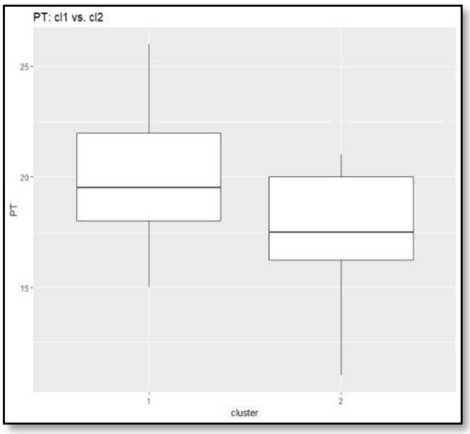


Figure 4.18.16 PT in Cluster1 and Cluster2

Despite the attempt to create two equal clusters of 12 individuals, the analysis creates a cl1 composed of 16 individuals and cl2 composed of 8 individuals. This suggests that cl2 is actually formed by individuals overall displaying larger differences with the mean of the population, and it is possible to expect, therefore, that the EEG waveform of cl2 reflects the aggregation of the effects found in cl4 in the first analysis and cl3 in the second one: namely, I expect that cl2 displays a two-way modulations of the response to the 3 accents manipulation, and that it will be strongly guided by the presence of acoustical saliency.

4.19. ERP – Final Cluster

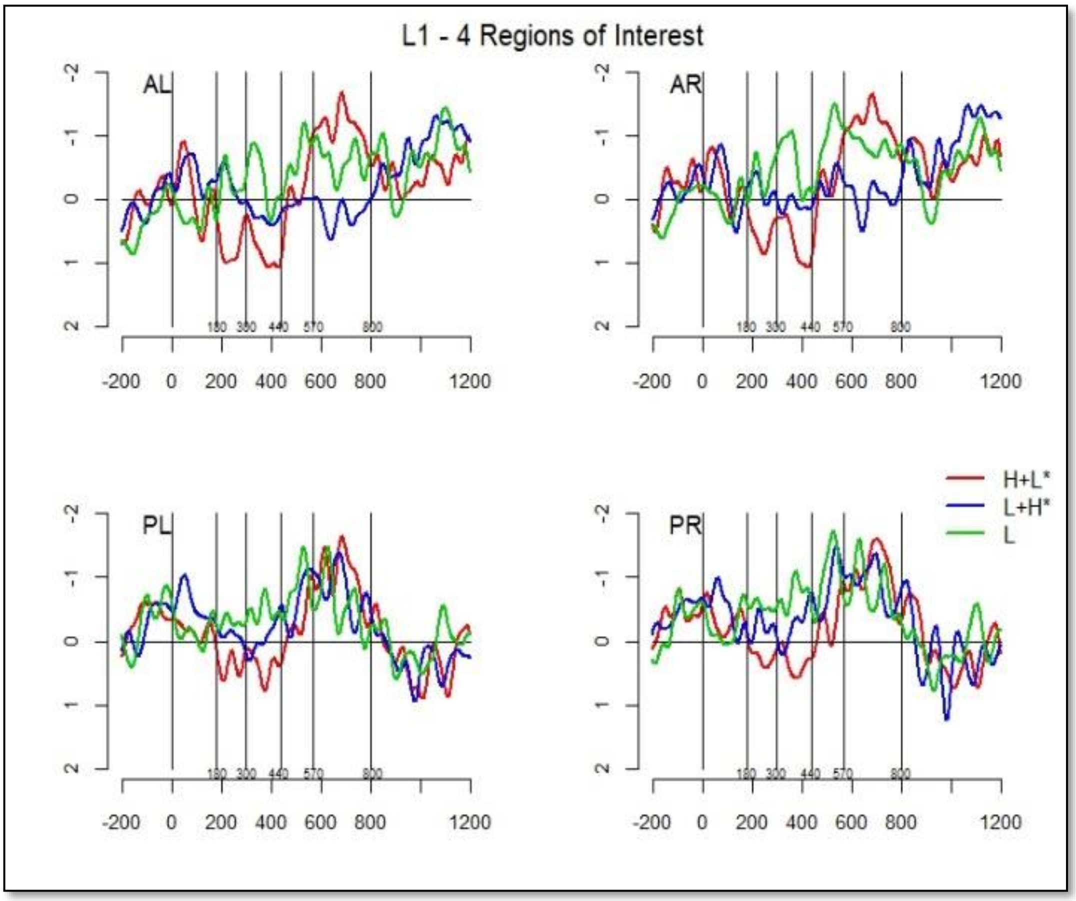


Figure 4.19.1 4 ROIs in L1 speakers for final comparisons

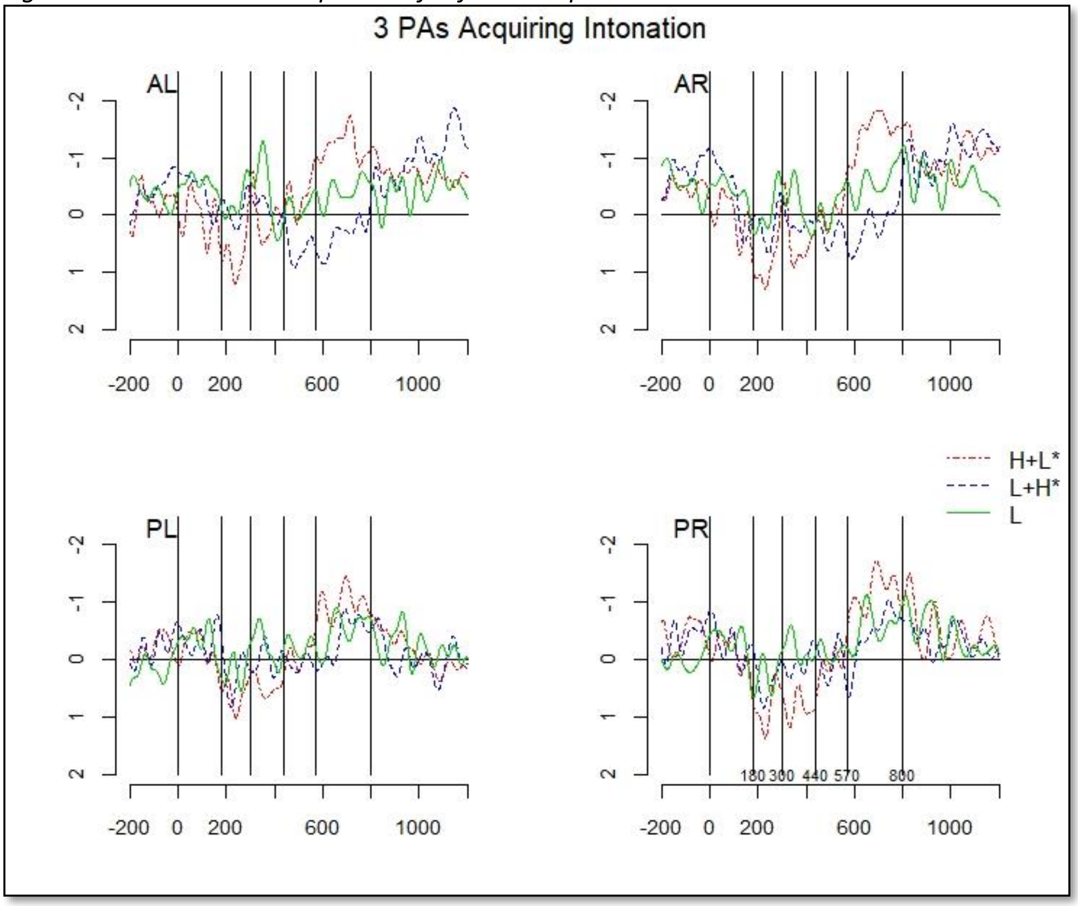


Figure 4.19.2 Final Comparison - 4 ROIs of Talented L2 learners

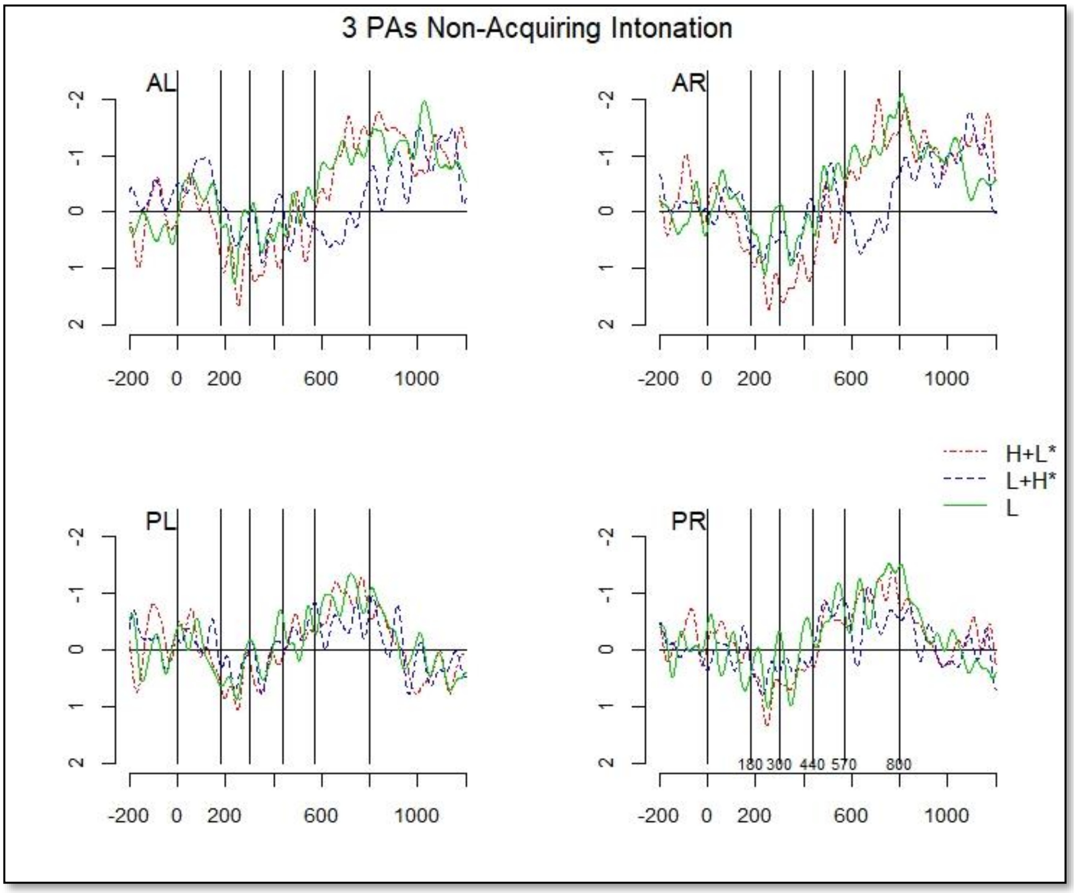


Figure 4.19.3 Final Comparison - 4 ROIs of Non-Talented L2 learners

4.19.1. Results of main factor PROSODY and lateral topographic distribution

4.19.1.1. Cl1 - Lateral

Over lateral electrodes, in the 180 – 300 ms time window, the main effect of PROSODY is present $F(2,30) = 8.54$; $p = 0.001$. In the 300 – 440 ms time window, the analysis indicates a main effect of PROSODY $F(2,30) = 6.07$; $p= 0.006$; and an effect of the topographic distribution $F(3,45) = 2.99$; $p = 0.04$. In the 570 – 800 ms time window, an effect for Prosody $F(2,30) = 19.28$; $p = 4.12$ e-06; and an interaction of the topographic factor with the main prosody effect are found $F(6,90) = 6.13$; $p = 2.30$ e-05.

4.19.1.2. Cl2 - Lateral

Over lateral electrodes, in the 180 – 300 ms time window, there is no significant effect.
In the 300 – 440 ms time window, only an effect of the topographic factor is present $F(3,21) = 3.55$; $p = 0.03$.
In the 570 – 800 ms time window, there is both an effect for Prosody $F(2,14) = 14.30$; $p = 0.0004$; and an interaction of the topographic factor with the main prosody effect $F(6,42) = 5.30$; $p = 0.0004$.

4.19.1.3. Pairwise Comparisons in 4 ROI for each cluster

Pairwise comparisons, FDR corrected for 3 within-subjects contrasts, over the lateral factor, show in cl1 the presence of an effect for PROSODY in the anterior regions in the 180-300 ms time window.
In the Anterior Left quarter there is only a tendency in 2 comparisons: Congruent vs. Deaccentuation t (15) = 2.23; p = 0.06; C.I.(0.04 -1.80); Congruent vs. Contrastive t(15) = 2.30; p = 0.06; C.I.(0.05 – 1.47);
In the Anterior Right quarter, the same 2 comparisons show a main effect: Congruent vs. Deaccentuation t (15) = 4.87; p = 0.0006; C.I.(0.52 – 1.32); Congruent vs. Contrastive t(15) = 2.73; p = 0.02; C.I.(0.15 – 1.18);
In the 330 – 440 ms time window: there is a significant comparison in the Anterior Right quarter: Congruent vs. Deaccentuation t (15) = 3.06; p = 0.02; C.I.(0.18 – 0.98); a tendency in the Posterior Left quarter in the same comparison: Congruent vs. Deaccentuation t (15) = 2.60; p = 0.06; C.I. (0.13 – 1.28); and 2 tendencies in the Posterior Right quarter in the Congruent vs. Deaccentuation comparison t(15) = 2.73; p = 0.05; C.I. (0.22 – 1.80) and in the Congruent vs. Contrastive comparison t(15) = 2.30; p = 0.05; C.I. (0.06 – 1.47).
In the 570 – 800 ms time window there is a strongly 3-way significant effect in all quarters apart from the Posterior Left quarter and a one tendency in the Posterior Right.
In the Anterior Left quarter all comparisons are significant: Congruent vs. Deaccentuation t(15) = -3.03; p = 0.01 C.I.(-1.35 - -0.24); Congruent vs. Contrastive t(15) = -4.26; p = 0.002 C.I. (-2.35 - -0.78) and Contrastive vs. Deaccentuation t(15) = -2.81; p = 0.01; C.I. (-1.37 - -0.19).
In the Anterior Right quarter as well, all comparisons are significant: Congruent vs. Deaccentuation t(15) = -3.28; p = 7.54 e-03; C.I. (-1.46 - -0.31); Congruent vs. Contrastive t(15) = -7.42; p = 6.444 e-06; C.I. (-2.13 . .1.18) and Contrastive vs. Deaccentuation t (15) = -2-28; p = 3.73 e-02; C.I. (-1.48 - -0.05).
In the Posterior Right quarter 2 comparisons are significant: Congruent vs. Deaccentuation t (15) = -3.44; p = 0.009; C.I. (-0.99 – 0.23); Congruent vs. Contrastive t (15) = -3.23; p = 0.008; C. I. (-1.24 - -0.25).

On the other hand, pairwise comparisons, FDR corrected for 3 within-subjects contrasts, over the lateral factor, show in cl2 the presence of only two tendencies in the Anterior Right quarter in the 180-300 ms time window in the Congruent vs. Deaccentuation comparison t (7) = 2.03; p = 0.14; C.I. (-0.11 1.48) and in the Congruent vs. Contrastive t (7) = 1.93; p = 0.14; C.I. (-0.11 - 1.15).
In the 300-440 ms time window there are only 3 tendencies right lateralized: in the Anterior Right quarter there are two tendencies in the Congruent vs. Deaccentuation comparison t(7) = 2.06; p =0.12; C.I. (-0.13 – 1.83) and Congruent vs. Contrastive comparison t(7) = 2.47; p = 0.12; C.I. (0.04 – 1.62); in Posterior Right quarter there is only a tendency for the Congruent vs. Deaccentuation comparison t (7) = 3.04; p = 0.06; C.I. (0.08 – 0.64).
Finally, in the 570-800 ms time window there are two significant comparisons both in the anterior region: in the Anterior Left quarter the Congruent vs. Contrastive comparison is significant t (7) = -4.04; p = 0.02; C.I. (-1.88 - -0.49) and the Contrastive vs. Deaccentuation one t (7) = -3.15) p = 0.02; C.I. (-2.06 - -0.29).
In the Anterior Right quarter, the same two comparisons are also significant: Congruent vs. Contrastive t (7) = -4.64; p = 0.007; C.I. (-2.95 – -0.63) and Contrastive vs. Deaccentuation t (7) = -3.57; p = 0.02; C.I. (-2.26 - -0.46).

4.20. *Discussion of the final cluster comparison in the L2 sample*

The final comparison between individuals displaying a Language-Aptitude for L2 prosodic processing and individuals lacking a particular talent for accent in this dataset highlights that: within the first cluster, cl1 are now grouped individuals that averagely began to acquire German as L2 at the age of 17. In opposition individuals in cl2 began at the mean age of 19. Although a direct comparison of mean differences is not advisable given the highly different number of participants in each cluster, the mean age does not appear to be largely different in the two groups. The variance pertaining to the AoA variable, nonetheless, shows to be largely different in the two groups. Individuals in cl2, thus, range from the lowest to the highest AoA values. The same observation is also sustainable for the values of the variable Age of 1st Immersion. Shifting from a mean age of 22 years old in cl1 to almost 25 in cl2 the difference appears not to be large, the variance values nonetheless indicate that cl2 comprehends the oldest individuals of the sample. An analogous but reversed effect is observable for the variable Att_Dett_30 (exceptional attention to details) that now shows to have equivalent mean values in the 2 clusters, but the individuals with the highest values in this variable belong now to cl1. The most polarized variables showing a trend to be consistently different from the two groups are then the two variables measuring the empathy construct: EC and PT, which show to averagely higher in cl1 compared to cl2.
The ERP analysis confirms the predicted results. The combination of the presence of a slightly lower AoA and Ao1st Immersion, with rather homogeneous and low variance, together with values of the variable measuring the exceptional attentional allocation to details, and generally consistently higher values in variables associated with the ToM construct such as the Empathy Concern and the ability to assume the interlocutor's perspective, together predict a rather advanced processing of L2 prosodic marking on the basis of information structure, modulated by both the hierarchical modulation of referential Givenness and the presence of a focus-background manipulation. In particular, the results highlight in cl1 a two-way frontal modulation of the auditive response component: P200, significant in the right anterior quarter and present as a tendency in the left anterior quarter and right posterior one. Moreover, cl1 displays a two-way modulation of the N400, larger for the Congruent vs. Deaccentuation comparison, and weakly emerging also in the comparisons with the contrastive accent. In final instance, cl1 displays a strong and stable 3-way modulation of the late-positivity component, that also emerges posteriorly.

On the other hand, cl2 only displays the presence of a tendency in the auditive discrimination of Congruent and Contrastive accents vs. Deaccentuation, present only in the right anterior quarter. It displays only a tendency in the N400 modulation emerging only in the right posterior quarter opposing the Congruent and Deaccentuation patterns, and mainly displays a two-way modulation of the late-positivity in components, emerging only in the anterior areas in correlation with the Contrastive accent condition. As previously reported, this two-way modulation can be presumably strongly due to the auditive processing of the final salient F0 peak in the contrastive condition.
Overall the analysis of L2 individual differences strongly supports the validity of the concept of Language Aptitude and promotes the investigation of individual predispositions in specific and separated areas of linguistic acquisition.

Concerning the acquisition of the L2 prosodic marking in function of the pragmatic structure, these results highly confirm the ongoing debate in the literature suggesting the strong connection between the auditive processing of pitch modulations and pragmatic linguistic and paralinguistic aspect of language.

The analysis of individual differences, pertaining the aspects of cognition that are bound to social communication and interaction and are captured by the construct of empathy and ToM, shows that these aspects of the single individual profiles are relevant for the auditive processing of suprasegmental features and the decoding of the discourse structure and pragmatic communicative intentions of the interlocutors. Also, the analysis indicates that the investigation of the sensory auditory processing of pitch can be a useful methodology to identify the presence of possible atypical development. Nonetheless, it has to be kept in mind that intonation shows a refined and detailed association with pragmatics; therefore, it is advisable to develop tools that precisely address the manipulation of specific intonation patterns and pragmatic functions, with the accuracy of excluding any other interfering aspect in the lexical material and phonetic implementation. As the study demonstrates, the presence of a steeper high F0 peak can change the pragmatic functional interpretation of the PA and, thus, its processing.

Specifically, by observing the variables classically manipulated in SLA studies, this study confirms the advantage in the acquisition of phonological and phonetic aspects of the L2 granted by an early acquisition. Investigating only late learners, overall reaching a comprehensive level of acquisition between the B2 and the C1 CEFR level, we can see that at least one third of them, characterized by higher AoA and shorter periods of training, has almost no auditive response to a 3-way pitch accents modulation and primarily uses, as intake for the following processing stages, the parameter of acoustical salience. Moreover, we can see that all male individuals belonging to the samples display a general difficulty in capturing the PA modulation.

Overall this suggests that L2 language teaching should enhance the training of phonological and phonetic aspects of language in the classes and should invest in the earliest formation of children of the youngest possible scholar age. Courses specifically training only adults can promote the institution of specific training classes specifically devoted to the acquisition of phonological aspects, stressing the importance of suprasegmental features within a linguistic system.

4.21. Conclusion

This study has investigated in L1 and L2 the auditive processing of prosodic marking in L1 and L2 learners of German. It has investigated the specific relationship that exists in German, and other West-Germanic languages (English and Dutch) between the alignment of specific PAs and the underlying discourse information structure.
The study has been designed initially to fully replicate the original study of Schumacher & Baumann (2010) addressing, in addition, the prosodic processing in L2. The original study manipulated the pitch variation in the experimental setting, referring to the theoretical concept of referential Givenness and its connection with the cognitive activation of a referent, that in German, is associated with 3 different PAs according to the Effort-Code proposed by Gussenhoven. Precisely Deaccentuation marks the presence of a fully activated Given referent, and a high central peak H* has been found in association with completely new and inactive referents. Accessible referents representing a stage of semi-activation have been found to be associated with the falling H+L* PA. The original paper systemically aligned the presented accent patterns always to an accessible semi-active referent standing in a whole-part relationship with the previously presented context sentence. Schumacher & Baumann (2010) found a 3-way modulation of the N400 signaling the higher integration cost of the processing of incongruent prosodic marking and a 2-way effect in the late-positivity component, functionally interpreted as the cognitive effort to introduce a new referent in the available

discourse representation. The 2-way effect indicated that the marking of a semi-active referent with an H* accent signaling the presence of a completely new referent does not generate a conflict with the processed information. On the contrary, Deaccentuation demands a higher effort since it indicates the presence of a textually or situationally given element that was not present in the manipulation. During the adaptation of the original material, in order to be suitable for L2 participants, with an increased acoustical intelligibility and a simplified lexicon, a pitch manipulation has been recorded, that in the occurrence of the original planned H* accent, displays a rapid and high rise, causing the effect to be possibly classified as L+H* and interpreted as a Narrow Focus accent.
The current experiment, therefore, introduced in one condition a focus-background manipulation that was not present in the original study.

The analysis of L1 results in comparison with Schumacher & Baumann (2010)'s results highlights that the slowing tempo enhanced the auditive response to the pitch manipulation in L1 participants, coherently with the Schumacher & Baumann (2010) results. This allowed to investigate the modulation of a P200 component. In addition, the N400 modulation has been fully replicated, showing that the change of PA did not overall affect the "economy" of the integration of prosodic cues with semantic and syntactical information, on one side, and on the other with the resolution of the bridging inference at the basis of the whole-part relationships and repartition of the discourse structure in function of both Givenness and Focus manipulation. Nonetheless, it is observable a strong anterior distribution of the effects that is mostly evident in the late-component time window. The late-positivity shows to reflect both the referential organization of Givenness and of the Focus-Background structure. Also, I have postulated that analogously to the emerging of the P200 component as auditive response, the presence of the highly acoustically salient peak, that it has been introduced, can modulate the late-positivity. These results support the previously found evidence of Chen et al., (2014) suggesting to separately investigate the modulation of Givenness and the modulation of Focus. The original intention was in line with Chen and colleagues, and the comparison between the results of Schumacher & Baumann's results with the L1 components coherently highlights that the introduction of a possible focus element enhances the late positivity response and that this positive response it is larger in the so-called Contrastive condition, compared to the Deaccentuation condition.

L1 vs. L2 analysis was primarily guided by the SLA literature indicating the prosodic marking of pragmatic aspects in German is a more marked relationship to acquire compared to the rules guiding the prosodic marking in the native language of the L2 speakers: Italian. The ongoing debate, in this respect, controversially reports the lack of Deaccentuation in the L1 Italian prosodic system in function of Givenness and the presence of recognizably different prosodic marking in function of focus type: broad vs. narrow informative vs. narrow contrastive. In particular, the classical literature, on typological differences between the prosodic marking in Germanic and Romance languages, indicated that Romance languages disfavor Deaccentuation when the pragmatic organization of discourse requires to shift the prosodic prominence on other referents in the sentence, compared to the default metrical distribution. On the contrary, Romance languages seem to favor a syntactical reorganization of the sentence guided by the presence of a focus element positioned in sentence-final position, matching the metrical structure. The monolithic division between Romance and Germanic languages, concerning non-plastic vs. plastic languages, has been questioned, suggesting the presence of relevant differences, also among Romance languages, that indicates the presence of a continuum of plasticity also among Romance languages. In this respect, Italian shows to reflect a higher level of plasticity compared for example to Spanish. Despite the ongoing debate on plasticity and the presence of controversial results, the lack of Deaccentuation in Italian has been reported, as well

as the resistance to acquire the manipulation of the Deaccentuation contour associated to Given referents in Germanic L2s. None of the available studies, nonetheless, investigated the cognitive processing of Deaccentuation of Italian native speakers, and L2 learners of a Germanic language, employing neurophysiological measures such as ERPs.
Our results of group comparison show to be in line with the classical position comparing plastic and non-plastic language since Italian participants averagely show to be sensible to the presence of the congruent PA corresponding in German to the semi-active state of a referent and in Italian to the presence of broad focus accent. In addition, they show a rather similar sensitivity, compared to L1 speakers, to the presence of the exceptionally high peak, introduced in the Contrastive Focus manipulation, displaying a 3-way modulation in the late-components. On the contrary, they do not auditorily attend, neither subsequently consider in the integration of the discourse model, the pitch modulation corresponding to Deaccentuation. Looking more deeply in the differences between L1 and L2 processing, the results support the correlation between the presence of narrow focused unit with the increase attentional resources allocation, and supports therefore, the theoretical position indicating focus as a unit helping the processing of information, directing the attention of the hearer towards what it is assumed to be informative in the conversation.
Moreover, the study highlights that the attentional mechanisms are strongly relevant in the acquisitional stage of a language because they show to modulate the step from the processing of the input signal to the generation of the intake. It is the intake; therefore, that determines what will be taken into consideration in the later stages.

In final instance, the experiment supports the presence of a tight connection between prosodic marking and the decoding of the pragmatic structures since it indicates that both, referential Givenness and relational communicative intentions of the interlocutors, predict the modulation of PA and their interpretation. This is strongly highlighted by the analysis of individual differences, showing that individual variation in the presence of empathic and Theory of Mind traits, as well as of traits belonging to ASC, modulates the individual aptitude in L2 prosodic processing, namely: traits enhancing the decoding of pragmatic aspects of communication also enhance the sensibility to PAs modulation in L2, and traits correlating with pragmatic impairments are associated with a poorer L2 processing of suprasegmental features.
In final instance, the study supports the advantage of the earlier acquisition in life of L2 compared to late acquisition, for what specifically belongs to the phonological and phonetic level of a language, as well it suggests a marginal advantage of female learners over male learners.
The limits of the study are mostly related to the presence of an experimental condition possibly manipulating two pragmatic levels of interpretations: referential and relational, that was not originally planned in the design. Nonetheless, the presence of this spurious condition has helped to disentangle the processing of Givenness and Focus mainly highlighted by the late-positivity component. Further research, precisely separately addressing the two Givenness and Focus manipulation in L2, is needed to further clarify which processing aspects are shared, and which ones specifically belong to the two separate dimensions in L2.
Overall the study introduces relevant information in the ongoing debate in SLA of suprasegmental features.

Summary and Conclusion of the thesis

This thesis has investigated the perception of intonational categories in adult native and second language speakers. Experiment 1 and Experiment 2 in Chapter 2 and Chapter 3 have investigated behaviorally, at the low sensory psychophysical level, the categorization process of final BTs in function of the recognizability of segments. Experiment 3 in Chapter 4, by employing the EEG technique, addressed in L1 and L2 the integration of prosody and information structure.

One aim of the thesis was to link evidence coming from independent fields of research that along the years contributed to signal the complexity of the subject.

The idea of linking the time course of the processing stages as well different scientific areas can be considered innovative and be adopted as a baseline for the development of a specific modeling of pitch processing, considering the implications coming from more than one theoretical framework. The innovation does not consist here in the introduction of a completely brand-new technique but in the unification of the evidence coming from different perspectives. Such an approach, I believe, can overcome the limits given by the adoption of a single paradigm and theoretical framework. Moreover, it encourages the exchange of expertise coming from the area of theoretical and acquisitional linguistics with the expertise focusing on signal processing, and the expertise coming from the clinical studies investigating atypical populations. Although the interdisciplinarity of the area is well recognized, the specific skills, terminology, and goals developed in the independent disciplines may not be commonly shared. This can stop the developing of a more comprehensive theory of pitch processing. This work on one side predicted a strong interplay between the auditory level, the discourse modeling, and the individual traits, and within the single studies made specific predictions about the manipulation of the single variables. On the other hand, exploratively addressed the possibility to integrate the different measures within a common framework. The unification of different approaches entails a starting difficulty establishing a common ground, axioms, and terminology. Moreover, it entails a more complex statistical modeling of the results, that correctly allows to observe the presence of relevant correlations. With this respect, in the absence of replicated results, I consider it explorative, meaning that future research will have to confirm these results and refine the links and relationships between each aspect of the multidisciplinary approach.

In particular, the theoretical debate introduced in Chapter 1 highlighted how the notion of intonational category is a key point for all the following processing stages, and for the application of a theoretical model, as in the case of the AM theory, to the field of SLA, and to the investigation of different populations. In presence of two distinct theoretical approaches: on one side, there is the AM that structurally decomposes the tune in discrete abstract representations, and further specifies them by the phonetic implementation; on the other side, there is the KIM model, and models with a comprehensive speech perspective such as the Polysp model, strongly supporting the view that specific phonetic features are relevant for the processing of pitch (and speech) and overcome the necessity of a phonological representation.

I believe it is beyond the scope of the thesis to discuss the role of phonological representations comprehensively. On the other hand, I find that there is a significant amount of evidence signaling the importance of the time-domain, in terms of duration, and in terms of alignment properties, to keep considering it only a realizational phonetic dimension, as it is for example considered now in the field of SLA.

The second aspect that emerges from the literature is that the rate of variation of a pitch contour is also evaluated within the categorization process. The indication comes jointly from the research on alignment and the neuroscientific evidence. Faster and slower transitions show to be processed through different networks, indicating that the fastest spectral transitions, present in segments, are the ones more categorically perceived and their processing appears to be left lateralized. On the other hand, slower transitions, such as the modulation of pitch, show to be less categorically perceived and their processing is more right lateralized. The presence of a stable and trained language system plays a role in the categorization process, and the recognition of pitch categories since evidence reports native speakers of tonal languages perceiving tones more categorically than speakers of intonational languages.

Language, and speech, presenting a signal that combines overall fast and slow transitions integrating pitch movements and segments, can switch the perception modality, from non-linguistic to linguistic pitch perception. Linguistic pitch perception also showed to be left lateralized.

Language through the presence of segments provides anchor points for the pitch movements. The anchor points have phonetic and acoustical features to which during perception is allocated attention. The presence of acoustically salient modulations has to be detected within stimuli of longer durations when compared to segments, and its categorization is likely to occur through extrinsic reference. This implies the storage in memory of longer parts of speech and the evaluation of a signal that varies over time. The combination of these phenomena decreases the presence of Categorical Perception as classically described in consonants transitions. Nonetheless, language shows not to be a prerequisite necessary in order to have CP. Categories can be formed through the extrinsic evaluation of several incoming stimuli, by detecting at least a common property present across stimuli that can be used as a criterion for the formation of a new category. In the case of pitch, the evidence indicates that the neural networks involved in the presence or absence of linguistic information are different. Thus, the decoding of recognizable linguistic features guides the activation of the most efficient perceptual strategy; in adjunct, it activates the categories available in the specific language system and generates an expectation on the income stimuli.

In speech, and during the perception of linguistic pitch-modulation, the perceived auditory category from this stage on then has to be integrated with the content of the currently processed information, and the content of the previously processed information. At this stage the information deriving from the context composed of the discourse model, (and potentially also by non-verbal cues present in the scene, such as facial expressions, in a naturalistic setting) and by an accessible semantic representation of the currently processed word has to be available and has to match the auditory category, following the linguistic structure of the referring linguistic system.

The evidence coming from the 3 experiments in this work supports the view that in the absence of congruent information coming from these 3 domains, the mapping between the auditory information and the pragmatic interpretation is not likely to efficiently occur. Also, the thesis supports the view, that in cases where the linguistic structure is coded differently in L2, compared to native language, the categories and the mapping rules of the native language interfere with the processing in L2. The interference can be potentially controlled by the presence of attentional and mnemonic resources at the individual level, and by the presence of a prolonged training in L2 started within adolescence.

In Chapter 2, the integration of segmental, semantic, and pragmatic information with the auditory stimuli continuously shifting between two F0 ceiling values showed to affect the emerging and recognizability of two pragmatic interpretations: question vs. statement. The categorization process shifting from the processing of highly recognizable linguistic stimuli, to lowly recognizable linguistic stimuli, reflected in the mathematical parameters of a logistic curve – slope and midpoint - has been modelled through Linear Mixed Models in two directions, highlighting that the presence of segments changes the perceptual strategy and forces the classical paradigm to switch from simply testing Categorical Perception, to allowing the investigation of a categorization process that can be independent of the linguistic nature of the stimuli. After splitting the non-linguistic modality from the linguistic modality, the presence of segments modulated by the recognizability factor showed a trend to modulate a linear relationship with the recognition of two categories: statement and question.

The modeling of individual pragmatic variability showed to moderately correlated with the categorization process of linguistic pitch categories. In particular, the model shows to fit the categorization process starting from stimuli with poorer information that linearly increases across condition (sequence 2). This indicates, in a broader perspective, for example, that language learners exposed to a new, poorly recognizable, spoken language system could show an advantage during the acquisitional process when displaying a specific set of cognitive resources.

Moreover, it signals that the separated fields investigating cognitive deficits, impairments, and atypical developments, on one side, and normo-typical language acquisition on the other one, can be linked, and simultaneously make use of the opposite field's results, through the evidence coming from the research on linguistic perception and processing of pitch movement, since the experiment supports the presence of a common mechanism at the very early stage and the presence of a strongly different top-activation instantiated by language experience.

Data on cognitive impairments can be used to advance in the linguistic description of the intonational structure, and of the abstract cognitive properties of intonational categories, as well as, in the opposite direction, a modeled intonational structure could improve the testing and diagnosis criteria of atypical linguistic behaviors in pathological populations.

In Chapter 3, the full application of the CP paradigm in a linguistic and a non-linguistic condition, through the adoption of a new adaptive procedure highlighted that the perception of F0 is affected by the overall frequency range, advantaging the perception of lower ranges of frequencies. This was particularly evident in the perception task performed with linguistic material compared to the perception task performed in the humming condition.

These results supported the results provided by the several experiments and theoretical model of David House, and supported the motion proposed by some authors within the field of Categorical Perception of intonation to consider the falling declining F0 movement in statement , the expected non-marked F0 movement in speech (Gussenhoven, 2012; Ladd, 1996; Schneider, 2012; Van Heuven & Kirsner, 2004).

The absence of the evidence of pure CP, but the presence of a correlation between increasingly steep slopes within the identification task, with decreasingly discrimination thresholds highlight that the two intonational categories have been recognized and made available to the cognitive systems of the participants, but that probably the physical properties of the signal play a stronger role in the discrimination task compared to the identification one.

This indicates that the recognition of an intonational category is a process involving higher linguistic representations, while the discrimination between two pitch movements from pitch height is probably performed by low sensory level mechanisms.

In line with this explanation of the involved cognitive processes, the pragmatic profiles of the participants do not significantly correlate with their discrimination sensitivity since the advantage of efficient, pragmatic decoding might largely impact in the steps following the discrimination process.

In Chapter 4, the EEG experiment provided relevant evidence for the debate on the processing mechanisms pertaining to the representation of discourse structure.

By replicating Schumacher & Baumann (2010)'s paradigm and by varying the stimuli including a possible Contrastive focus modulation within a modulation centered on referential givenness, the experiment is able to replicate and confirm the previous results indicating the association of the N400 component with the integration of the semantic properties of the referents within the current state activation of discourse.

Despite the recording of new stimuli, with a simpler choice of vocabulary, and despite the change of the voice from male to female, and of the speech tempo (slower than in the original study), the data of native speakers of German replicated the graded modulation of the N400 in function of referential Givenness. This indicates that interpretation of the N400 responses in experimental settings shifting from the visual modality to the auditive modality stably holds, and it is a useful parameter in studies comparing native and non-native perception.

On the other hand, the introduction of an extremely salient peak, possibly assuming a contrastive interpretation, highlighted the presence of a positive shift, emerging in the late-time window strongly responding to this specific condition.

The presence of speech produced at a slower tempo and marking the presence of peaks and valleys allowed to observe the modulation of early components, presumably and primarily responding to auditive attentional mechanisms, and reflected in the presence of a positive peak at about 200 ms.

The slower tempo also appeared to affect the duration of each component, delaying the onset and significantly increasing the duration of the late positivity window.

In final instance, the introduction of a different modulation seemed in line with the results and proposal of Chen, Wang, & Yang, (2014) supporting a theoretical and empirical differentiation of the cognitive mechanisms related to the decoding of referential givenness and relational properties of discourse such as the focus element.

This represents a relevant result also for the literature on cross-linguistic modeling of intonation in Romance and West-Germanic languages and for the debate on the nature of plasticity. In particular, the results related to a minimal variation of a phonetic parameter in the design, highlight that more precision is needed in the developing of experiments investigating acquisitional patterns of intonation related to discourse structure. More data are required to map the typological difference between and within Romance languages, opposed to West-Germanic ones.

The data concerning the perception on L2, highlight that the rarely occurring Deaccentuation pattern in Italian L1, as well as the presence of a phonological system, assumed to be strongly guided by metrical aspects and relational properties of discourse affect the way L2 learners perceive and process accessible information in German.

In particular, L2 learners show to have a binary response in the modulation of the N400, strongly based on the evaluation of the most congruent accentuation pattern (H+L*) that in Italian is not reported to be associated with the degree of accessibility of a referent, but rather with the presence of Broad Focus sentences. Also, at the group-level, the language learners show a binary response in the late positivity window in function of the presence of the Contrastive condition, associated with acoustical salience.

Overall at the group -level, L2 learners do not seem to adopt the evaluation of the degree of cognitive activation of a referent as a criterion for a congruent integration of the auditory information with the information previously processed in the discourse model.

With this respect, Deaccentuation does not appear to play a relevant role for L2 learners of German, native speakers of Italian.

On the contrary, the strongest effects within the ERP-response-patterns are related to the presence of a perceivable pitch movement, landing at a recognizable point in time within the syllable, that appears to be related to auditory selective attentional mechanisms.

On one side, this evidence indirectly supports the literature invoking a clear-cut division between plastic and non-plastic languages. On the basis of the reported attentional response, it can be possible to support the view that the intonational structure of Italian is primarily guided the presence of focus elements, emerging at the sensory level with the presence of salient pitch modulation. The data presented in Chapter 4, show that L2 learners primarily respond to the detection of acoustical prominence.

The presence of an attentional mechanism auditorily responding to prominence is in line with the literature presented in the introduction indicating a universal mechanism detecting rising contours, as deviating from a non-marked declination pattern. Moreover, it is also in line with the indication that the universal function of intonation interacts with the presence of a structurally organized system, that differently maps sound and discourse across languages. In absence of a clear mapping of Deaccentuation and Given information, a lack of acoustical prominence appeared to be not present within the attentional template of Italian L1 speakers, being harder to notice within the whole pitch contour, and thus not expected on the basis of the previous information. This leads to the fact that is not evaluated as non-congruent when matched with accessible information.

The presence of a non-congruent acoustically salient peak, as in the contrastive condition, is available in the Italian system, is not predicted by the discourse model modulating accessibility, but it is detectable during the online processing. Thus, the incongruency with the presence of accessible information can be processed.

On the other side, investigating the individual variability of both external (related to the training of the L2) and internal (related to the pragmatic characteristics of the participants) factors, the data suggest that the way single individuals adopt selective attentional mechanisms, and the degree of efficiency of their pragmatic interpretation of speech acts can impact on the acquisitional patterns.

Namely, the results of the cluster analysis investigating the role of covariate measures show that individuals with: i) higher empathic traits and ii) that can assume the interlocutor perspective, and that iii) as well display a stronger tendency in devoting a higher degree of attention to auditory processing, can overcome the intonational structure of Italian. These participants clearly show an almost native-like modulation of the N400 component on the basis of the degrees of referential givenness.

In final instance, the experiment indicates that the Age of Acquisition is a relevant factor in the acquisition of an L2, supporting the well-known relationship between the start of acquisition at a younger age and the higher probability to achieve L2 native-like processing.

As a final remark, it is desirable that future research on pitch processing will make use of a more comprehensive theoretical perspective that overcomes the constraints of the adoption of a single description, and instead considers both the plurality of supporting evidence, as well as the problematic aspects coming from different frameworks. In particular, it is desirable that both the theoretical linguistic descriptions and the L2 acquisitional models, merge with the cognitive models of perception and the neurophysiological evidence providing both time and space resolution of the activation of specific neural networks. Moreover, it is desirable that psycholinguistic and neurolinguistic research evidence merges with the clinical evidence that would ultimately benefit from a comprehensive approach, improving the diagnostic criteria and the possible therapeutic intervention.

The quantity of information that pitch carries within normal conversations, and outside the domain of language, indicates that it serves a fundamental cognitive function, possibly linking different networks: the verbal sphere, as well as the expression of emotion, the processing of music, and the recognition of human beings, animals, and specific objects like musical instruments.

The understanding of its functioning in cognitive and neurophysiological terms could be a possible advantage for each of the mentioned areas of interest. The advancing of the knowledge we have on pitch and auditory processing, its integration with cues coming from different modalities, (such as the facial expressions or the posture and gestures), as well as its integration with higher order linguistic structures pertaining the lexicon and grammar, can represent a step forward also for the understanding of the atypical functioning of this combination of features. The developing of better models, paradigms, and diagnostic criteria can ultimately improve the quality of people's life.

Moreover, language is the widely adopted system used by millions of individuals to communicate, make choices, express feelings, work, traveling, or moving across continents for the most varied reasons – pleasure and necessity. Refined knowledge of the processing of pitch in language, and ultimately of the speech processing and language acquisition, can improve the teaching strategies and the educational systems. Knowing about the individual variability of different people demonstrating either an exceptional talent or experiencing difficulties and frustration in the acquisition of spoken languages can lead to the improvement of educational strategies fitted on the single person needs, based on their ages, their attitudes, and their previous experiences.

Last but not least, nowadays pitch within speech is modeled and implemented in the electronic technologies, and it is used to achieve a high standard of automatization. A better understanding of the role of pitch in the expression of referential and relational aspects of discourse, in the expression of explicit and implicit aspects of verbal communication, can improve the developing of the current technologies, such as the text-to-speech and the voice recognition systems. The processing of voice features is linked nowadays to a plurality of private information. Voice is used in instant messages technologies, to access bank accounts and private resources. A more efficient modeling can increase the efficiency and security of the disposable devices that we use every day.

I believe that a truly multidisciplinary perspective could address and achieve remarkable results in each of the mentioned areas. This work is my step towards that direction.

References

Abrahamsson, N., & Hyltenstam, K. (2008). The robustness of aptitude effects in near-native second language acquisition. *Studies in Second Language Acquisition*, *30*(4), 481–509.

Abramson, A. S. (1978). Static and dynamic acoustic cues in distinctive tones. *Language and Speech*, *21*(4), 319–325.

Abramson, A. S., & Lisker, L. (1970). Discriminability along the voicing continuum: Cross-language tests. In *Proceedings of the sixth international congress of phonetic sciences* (Vol. 196, pp. 569–573). Academia Prague.

Adams, C., & Munro, R. R. (1978). In search of the acoustic correlates of stress: fundamental frequency, amplitude, and duration in the connected utterance of some native and non-native speakers of English. *Phonetica*, *35*(3), 125–156.

Ahmed, F. S., & Miller, L. S. (2011). Executive function mechanisms of theory of mind. *Journal of Autism and Developmental Disorders*, *41*(5), 667–678.

Alain, C., & Arnott, S. R. (2000). Selectively attending to auditory objects. *Front. Biosci*, *5*, D202–D212.

Ames, H. (1969). Italian intonation patterns in conversation and reading. *Lingua e Stile*, *4*, 43–52.

Ammann, H. (1928). *Die menschliche Rede*. M. Schauenburg.

Amunts, K., Schleicher, A., & Zilles, K. (2004). Outstanding language competence and cytoarchitecture in Broca's speech region. *Brain and Language*, *89*(2), 348–353.

Anderson, J. (1987). The markedness differential hypothesis and syllable structure difficulty. *Interlanguage Phonology: The Acquisition of a Second Language Sound System*, 279–291.

Anderson-Hsieh, J., Johnson, R., & Koehler, K. (1992). The relationship between native speaker judgments of nonnative pronunciation and deviance in segments, prosody, and syllable structure. *Language Learning*, *42*(4), 529–555.

Arvaniti, A., Ladd, D. R., & Mennen, I. (2006). Phonetic effects of focus and "tonal crowding" in intonation: Evidence from Greek polar questions. *Speech Communication*, (48(6)), 667–696.

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychol. Sci.*, *9*, 321–324. https://doi.org/10.1111/1467-9280.00063

Atterer, M., & Ladd, D. R. (2004). On the phonetics and phonology of "segmental anchoring" of F0: evidence from German. *Journal of Phonetics*, *32*(2), 177–197.

Avesani, C. (1990). A contribution to the synthesis of Italian intonation. In *ICSLP* (Vol. 90, pp. 834–836).

Avesani, C., Bocci, G., Vayra, M., Zappoli, A., & Chini, M. (2015). Prosody and information status in Italian and German L2 intonation. In *Il parlato in [italiano] L2: aspetti pragmatici e prosodici /[Italian] L2 Spoken Discourse: Pragmatic and Prosodic Aspects,* (pp. 93–116). Milano: Franco Angeli.

Avesani, C., & Vayra, M. (2004). Focus ristretto e focus contrastivo in italiano. *Il Parlato Italiano. Atti Del Convegno Nazionale, Napoli*, 1–20.

Ayotte, J., Peretz, I., & Hyde, K. I. (2002). Congenital amusia: A group study of adults afflicted with a music-specific disorder. *Brain*, *125*(2), 238–251.

Backman, N. (1979). Intonation errors in second-language pronunciation of eight Spanish-speaking adults learning English. *Interlanguage Studies Bulletin*, 239–265.

Baddeley, A. D. (1992). Working memory. *Science*, *255*(5044), 556–559.

Baddeley, A. D., & Hitch, G. (1974). Working memory. *Psychology of Learning and Motivation*, *8*, 47–89.

Baron-Cohen, S. (1997). *Mindblindness: An essay on autism and theory of mind*. MIT press.

Baron-Cohen, S., Tager-Flusberg, H., & Cohen, D. J. (1994). *Understanding other minds: Perspectives from autism*. New York, NY, US: Oxford University Press.

Barsalou, L. W. (2005). Situated conceptualization. In *Handbook of categorization in cognitive science* (pp. 619–650). Elsevier.

Bastian, J., & Abramson, A. S. (1962). Identification and discrimination of phonemic vowel duration. *The Journal of the Acoustical Society of America*, *34*(5), 743–744.

Bastian, J., Eimas, P. D., & Liberman, A. M. (1961). Identification and discrimination of a phonemic contrast induced by silent interval. *The Journal of the Acoustical Society of America*, *33*(6), 842–842.

Baumann, S. (2005). Degrees of givenness and their prosodic marking. In *International Symposium on" Discourse and Prosody as a complex interface*.

Baumann, S. (2006). Information structure and prosody: Linguistic categories for spoken language annotation. *Methods in Empirical Prosody Research*, 153–180.

Baumann, S., & Grice, M. (2006). The intonation of accessibility. *Journal of Pragmatics*, *38*(10), 1636–1657.

Baumann, S., & Hadelich, K. (2003). Accent type and givenness: an experiment with auditory and visual priming. *Proceedings of the 15th ICPhS, Barcelona, Spain*, 1811–1814.

Baumann, S., & Schumacher, P. B. (2012). (De-)Accentuation and the Processing of Information Status: Evidence from Event-Related Brain Potentials. *Language and Speech*, *55*(3), 361–381.

Beale, J. M., & Keil, F. C. (1995). Categorical effects in the perception of faces. *Cognition*, *57*(3), 217–239.

Beckman, M. E. (1986). *Stress and non-stress accent* (Vol. 7). Walter de Gruyter.

Beckman, M. E., & Pierrehumbert, J. B. (1986). Intonational structure in Japanese and English. *Phonology*, *3*, 255–309.

Behnke, K. (1998). *“The acquisition of phonetic categories in young infants: a self organising neural network approach”. Doctoral Dissertation*. University of Nijmegen.

Best, C. (1995). *A direct-realist view of cross-language perception. Speech perception and linguistic experience: issues in cross-language research, ed. by W. Strange, 171–204*. Baltimore: York Press.

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege*, *1334*.

Bialystok, E. (1997). The structure of age: In search of barriers to second language acquisition. *Second Language Research*, *13*(2), 116–137.

Bialystok, E., & Hakuta, K. (1994). *In other words: The science and psychology of second-language acquisition*. Basic Books.

Bialystok, E., & Hakuta, K. (1999). Confounded age: Linguistic and cognitive factors in age differences for second language acquisition. *Second Language Acquisition and the Critical Period Hypothesis*, 161–181.

Birch, S., & Rayner, K. (2010). Effects of syntactic prominence on eye movements during reading. *Memory & Cognition*, *38*(6), 740–752.

Birdsong, D. (1992). Ultimate attainment in second language acquisition. *Language*, 706–755.

Birdsong, D. (1999). *Second language acquisition and the critical period hypothesis*. Routledge.

Birdsong, D. (2006). Age and second language acquisition and processing: A selective overview. *Language Learning*, *56*, 9–49.

Birdsong, D. (2007). Nativelike pronunciation among late learners of French as a second language. *Language Experience in Second Language Speech Learning*, 99–116.

Bley-Vroman, R. (1988). The fundamental character of foreign language learning. *Grammar and Second Language Teaching: A Book of Readings*, 19–30.

Boersma, P. (2001). Praat, a system for doing phonetics by computer. Glot International.

Bolinger, D. (1964). Intonation as a universal. In *Proceedings of the 5th Congress of Phonetics, Cambridge 1962* (pp. 833–848).

Bolinger, D. (1978). Intonation across languages. *Universals of Human Language*, *2*, 471–524.

Bolinger, D. (1989). *Intonation and its uses: Melody in grammar and discourse*. Stanford University Press.

Bongaerts, T. (1999). Ultimate attainment in L2 pronunciation: The case of very advanced late L2 learners. *Second Language Acquisition and the Critical Period Hypothesis*, 133–159.

Bongaerts, T., Planken, B., & Schils, E. (1995). Can late starters attain a native accent in a foreign language? A test of the critical period hypothesis. *The Age Factor in Second Language Acquisition*, 30–50.

Bongaerts, T., Van Summeren, C., Planken, B., & Schils, E. (1997). Age and ultimate attainment in the pronunciation of a foreign language. *Studies in Second Language Acquisition*, *19*(4), 447–465.

Bornkessel, I., Schlesewsky, M., & Friederici, A. D. (2003). Eliciting thematic reanalysis effects: The role of syntax-independent information during parsing. *Language and Cognitive Processes*, *18*(3), 269–298.

Boster, J. S. (2017). Categories and cognitive anthropology. In *Handbook of Categorization in Cognitive Science (Second Edition)* (pp. 75–106). Elsevier.

Bradford, E. E., Jentzsch, I., & Gomez, J.-C. (2015). From self to social cognition: theory of mind mechanisms and their relation to executive functioning. *Cognition*, *138*, 21–34.

Bruce, G. (1977). Swedish accents in sentence perspective. *Working Papers in Linguistics*, *12*(1).

Bruce, G., & Gårding, E. (1978). A prosodic typology for Swedish dialects. *Working Papers in Linguistics*, *17*(1).

Buddenhagen, R. (1971). Establishing vocalizations in mute mongoloid children. Champaign. *Ill.: Research Press, I97L*.

Büring, D. (2012). Focus and intonation. *Routledge Companion to the Philosophy of Language. London: Routledge*, 103–115.

Burkhardt, P. (2006). Inferential bridging relations reveal distinct neural mechanisms: Evidence from event-related brain potentials. *Brain and Language*, *98*(2), 159–168.

Burkhardt, P. (2007). The P600 reflects cost of new information in discourse memory. *Neuroreport*, *18*(17), 1851–1854.

Carlisle, R. S. (1991). The influence of environment on vowel epenthesis in Spanish/English interphonology. *Applied Linguistics*, *12*(1), 76–95.

Carretié, L., Mercado, F., Tapia, M., & Hinojosa, J. A. (2001). Emotion, attention, and the 'negativity bias', studied through event-related potentials. *International Journal of Psychophysiology*, *41*(1), 75–85.

Carroll, J.B., & Sapon, S. M. (1959). *Modern language aptitude test* (Psychological Corporation.). San Antonio, TX (USA).

Carroll, John B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge University Press.

Chafe, W. (1974). *Language and consciousness*.

Chafe, W. (1976). *Givenness, contrastiveness, definiteness, subjects and topics. C. Li (ed.), Subject and topic, 25-56*. New York, Academic Press.

Chafe, W. (1994). *Discourse, consciousness, and time: The flow and displacement of conscious experience in speaking and writing*. University of Chicago Press.

Chaudron, C. (1985). Intake: On models and methods for discovering learners' processing of input (Vol. 1, pp. 1–14).

Chen, L., Wang, L., & Yang, Y. (2014). Distinguish between focus and newness: an ERP study. *Journal of Neurolinguistics*, *31*, 28–41.

Chomsky, N. (1971). Deep structure, surface structure and semantic representations. *By DD Steinberg and LA Jakobovlts*, *183*, 216.

Chomsky, N. A. (1976). Reflections on language.

Chomsky, N., & Halle, M. (1968). The sound pattern of English.

Clark, H. H. (1977). Inferences in comprehension. *Basic Processes in Reading: Perception and Comprehension*, *2433263*.

Clark, H. H., & Haviland, S. E. (1977). Comprehension and the given-new contract. *Discourse Production and Comprehension. Discourse Processes: Advances in Research and Theory*, *1*, 1–40.

Cohen, H., & Lefebvre, C. (2005). *Handbook of categorization in cognitive science*. Elsevier.

Colantoni, L., & Steele, J. (2006). Native-like attainment in the L2 acquisition of Spanish stop-liquid clusters. In *Selected proceedings of the 7th Conference on the Acquisition of Spanish and Portuguese as First and Second Languages* (pp. 59–73). Cascadilla.

Conway, M. A. (2005). Memory and the self. *Journal of Memory and Language, 53*(4), 594–628.

Coppieters, R. (1987). Competence differences between native and near-native speakers. *Language*, 544–573.

Corder, S. P. (1967). The significance of learner's errors. In *IRAL-International Review of Applied Linguistics in Language Teaching* (pp. 161–170).

Cowles, H. W., Kluender, R., Kutas, M., & Polinsky, M. (2007). Violations of information structure: An electrophysiological study of answers to wh-questions. *Brain and Language*, *102*(3), 228–242.

Crowder, R. G., & Morton, J. (1969). Precategorical acoustic storage (PAS). *Perception & Psychophysics*, *5*(6), 365–373.

Cruttenden, A. (1981). Falls and rises: meanings and universals. *Journal of Linguistics*, *17*(1), 77–91.

Cruttenden, A. (1997). *Intonation* (Cambridge University Press).

Cruttenden, A. (2006). The de-accenting of given information: A cognitive universal? *Empirical Approaches To Language Typology*, *20*(8), 311.

Cruz-Ferreira, M. (1987). Non-native interpretive strategies for intonational meaning: An experimental study. *Sound Patterns in Second Language Acquisition*, 103–120.

Crystal, D. (1969). Prosodic Systems and Intonation in English. Cambridge Studies in Linguistics 1.

Cummins, F., Doherty, C., & Dilley, L. (2006). Phrase-final pitch discrimination in English. In *Proceedings of Speech Prosody* (pp. 467–470). TUDpress Verlag der Wissenschaften GmbH Dresden, Germany.

Curtiss, S. (1977). Genie: A psychological study of a modern-day "wild child.

Cutler, A., & Fodor, J. A. (1979). Semantic focus and sentence comprehension. *Cognition*, *7*(1), 49–59.

Cutting, J. E. (1982). Plucks and bows are categorically perceived, sometimes. *Perception & Psychophysics*, *31*(5), 462–476.

Cutting, J. E., & Rosner, B. S. (1974). Categories and boundaries in speech and music. *Attention, Perception, & Psychophysics*, *16*(3), 564–570.

Daneš, F. (1994). Involvement with language and in language. *Journal of Pragmatics*, *22*(3–4), 251–264.

De Leeuw, E., Mennen, I., & Scobbie, J. M. (2012). Singing a different tune in your native language: first language attrition of prosody. *International Journal of Bilingualism*, *16*(1), 101–116.

DeKeyser, R. (2017). Knowledge and skill in ISLA. In *The Routledge handbook of instructed second language acquisition* (pp. 15–32).

Diehl, J. J., Watson, D., Bennetto, L., McDonough, J., & Gunlogson, C. (2009). An acoustic analysis of prosody in high-functioning autism. *Applied Psycholinguistics*, *30*(3), 385–404.

Diller, K. C., & Walsh, T. M. (1981). " LIVING" AND" DEAD" LANGUAGES-A NEUROLINGUISTIC DISTINCTION. *Actes Du 5e Congrès de l'Association Internationale de Linguistique Appliquée*, *16*, 99.

Dilley, Laura C., & Brown, M. (2007). Effects of pitch range variation on f 0 extrema in an imitation task. *Journal of Phonetics*, *35*(4), 523–551.

Dilley, Laura Christine. (2005). *The phonetics and phonology of tonal systems*. Massachusetts Institute of Technology.

Dimitrova, D. V., Redeker, G., Egg, M., & Hoeks, J. C. (2008). Prosodic correlates of linguistic and extra-linguistic information in Dutch. In *CogSci 2008* (pp. 2191–2196). Cognitive Science Society.

Dimitrova, D. V., Stowe, L. A., Redeker, G., & Hoeks, J. C. (2012). Less is not more: Neural responses to missing and superfluous accents in context. *Journal of Cognitive Neuroscience*, *24*(12), 2400–2418.

D'Imperio, M. (1995). Timing differences between prenuclear and nuclear pitch accents in Italian. *The Journal of the Acoustical Society of America*, *98*(5), 2894–2894.

D'Imperio, M. (1996). Caratteristiche di timing degli accenti nucleari in parlato italiano letto. In *Atti del XXIV Convegno Nazionale dell'Associazione Italiana di Acustica* (pp. 55–60).

D'Imperio, M., & House, D. (1997). Perception of questions and statements in Neapolitan Italian. Presented at the Fifth European Conference on Speech Communication and Technology.

Dogil, G., & Reiterer, S. M. (2009). *Language talent and brain activity* (Vol. 1). Walter de Gruyter.

Dörnyei, Z. (2006). Individual differences in second language acquisition. In *AILA review* (Vol. 19, pp. 42–68).

Dryer, M. S. (2005). *Polar questions* (Vol. 116). Case.

Eales, M. J. (1993). Pragmatic impairments in adults with childhood diagnoses of autism or developmental receptive language disorder. *Journal of Autism and Developmental Disorders*, *23*(4), 593–617.

Eckert, H., & Laver, J. (1994). *Menschen und Ihre Stimmen. Beltz Psychologie Verlags Union*. Weinheim.

Eckman, F. R. (1977). Markedness and the contrastive analysis hypothesis. *Language Learning*, *27*(2), 315–330.

Eckman, F. R. (1991). The structural conformity hypothesis and the acquisition of consonant clusters in the interlanguage of ESL learners. *Studies in Second Language Acquisition*, *13*(1), 23–41.

Edelson, L., Grossman, R., & Tager-Flusberg, H. (2007). Emotional prosody in children and adolescents with autism. In *Poster session presented at the annual international meeting for Autism Research, Seattle, WA*.

Eimas, P. D. (1963). The relation between identification and discrimination along speech and non-speech continua. *Language and Speech*, *6*(4), 206–217.

Ellis, R. (2008). *The Study of Second Language Acquisition* (Oxford University).

Ellis, R., & Ellis, R. R. (1994). *The study of second language acquisition* (Oxford University.).

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science*, *11*(1), 19–23.

Eubank, L., & Gregg, K. R. (1999). Critical periods and (second) language acquisition: Divide et impera. *Second Language Acquisition and the Critical Period Hypothesis*, 65–99.

Falé, I., & Faria, I. H. (2006). Categorical perception of intonational contrasts in European Portuguese. In *Proceedings of Speech Prosody* (pp. 69–72). TUDpress Verlag der Wissenschaften GmbH Dresden, Germany.

Federmeier, K. D., & Kutas, M. (1999). A rose by any other name: Long-term memory structure and sentence processing. *Journal of Memory and Language*, *41*(4), 469–495.

Féry, C. (1993). *German intonational patterns* (Vol. 285). Walter de Gruyter.

Field, J. (2003). Promoting perception: Lexical segmentation in L2 listening. *ELT Journal*, *57*(4), 325–334.

Finn, A. S., & Kam, C. L. H. (2008). The curse of knowledge: First language knowledge impairs adult learners' use of novel statistics for word segmentation. *Cognition*, *108*(2), 477–499.

Firbas, J. (1964). On defining the theme in functional sentence analysis. *Linguistiques de Prague*, *1*, 267–280.

Flege, James E. (1995). Second language speech learning: Theory, findings, and problems. *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, *92*, 233–277.

Flege, James E., Schirru, C., & MacKay, I. R. (2003). Interaction between the native and second language phonetic subsystems. *Speech Communication*, *40*(4), 467–491.

Flege, James Emil. (1987). The production of "new" and "similar" phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics*, *15*(1), 47–65.

Flege, James Emil, & Port, R. (1981). Cross-language phonetic interference: Arabic to English. *Language and Speech*, *24*(2), 125–146.

Flege, James Emil, Yeni-Komshian, G. H., & Liu, S. (1999). Age constraints on second-language acquisition. *Journal of Memory and Language*, *41*(1), 78–104.

Fletcher, H. (1940). Auditory patterns. *Reviews of Modern Physics*, *12*(1), 47.

Flynn, S., & Manuel, S. (1991). Age-dependent effects in language acquisition: An evaluation of critical period hypotheses. *Point Counterpoint: Universal Grammar in the Second Language*, 117–145.

Fosnot, S. M., & Jun, S. (1999). Prosodic characteristics in children with stuttering or autism during reading and imitation. In *Proceedings of the 14th international congress of phonetic sciences* (pp. 1925–1928).

Fox, A. (1984). *German intonation: An outline*. Oxford University Press, USA.

Foxton, J. M., Dean, J. L., Gee, R., Peretz, I., & Griffiths, T. D. (2004). Characterization of deficits in pitch perception underlying 'tone deafness.' *Brain*, *127*(4), 801–810.

Francis, A. L., Ciocca, V., & Ng, B. K. C. (2003). On the (non) categorical perception of lexical tones. *Perception & Psychophysics*, *65*(7), 1029–1044.

Francis, A. L., & Nusbaum, H. C. (2002). Selective attention and the acquisition of new phonetic categories. *Journal of Experimental Psychology: Human Perception and Performance*, *28*(2), 349.

Frith, C. D., & Frith, U. (2006). The neural basis of mentalizing. *Neuron*, *50*(4), 531–534.

Frith, C., & Frith, U. (2005). Theory of mind. *Current Biology*, *15*(17), R644–R645.

Frith, U. (2003). *Autism: Explaining the enigma*. Oxford: Blackwell Publishing.

Frota, S. (2014). *Prosody and focus in European Portuguese: Phonological phrasing and intonation* (Routledge).

Fry, D. B., Abramson, A. S., Eimas, P. D., & Liberman, A. M. (1962). The identification and discrimination of synthetic vowels. *Language and Speech*, *5*(4), 171–189.

Fujisaki, H., & Kawashima, T. (1969). On the models and mechanisms of speech perception. *Annu. Rep. Eng. Res. Inst.*, *28*, 67–73.

Fujisaki, H., & Kawashima, T. (1970). Some experiments on speech perception and a model for the perceptual mechanism. *Annu. Rep. Eng. Res. Inst.(Fac. Eng., Univ. Tokyo)*, *29*, 207–214.

Fujisaki, H., & Kawashima, T. (1971). A model of the mechanisms for speech perception: Quantitative analysis of categorical effects in discrimination. *Annual Report of the Engineering Research Institute, Faculty of Engineering, University of Tokyo*, *30*, 59–68.

Gade, M., Druey, M. D., Souza, A. S., & Oberauer, K. (2014). Interference within and between declarative and procedural representations in working memory. *Journal of Memory and Language*, (76), 174–194.

Galantucci, B., Fowler, C. A., & Turvey, M. T. (2006). The motor theory of speech perception reviewed. *Psychonomic Bulletin & Review*, *13*(3), 361–377.

Gass, S.M. (1988). Integrating Research Areas: A Framework for Second Language Studies. *Applied Linguistics*, *9*(2), 198–217.

Gass, Susan M. (2017). *Input, interaction, and the second language learner*. Routledge.

Gathercole, S. E., Pickering, S. J., Knight, C., & Stegmann, Z. (2004). Working memory skills and educational attainment: Evidence from national curriculum assessments at 7 and 14 years of age. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *18*(1), 1–16.

Gibbon, D. (1998). German Intonation. In *Intonation Systems. A Survey of 20 Languages* (Cambridge University Press, pp. 78–95). Cambridge: Daniel Hirst & Albert di Cristo.

Gibson, E. J. (1969). Principles of perceptual learning and development.

Gibson, J. J. (1979). *The theory of affordances The Ecological Approach to Visual Perception (pp. 127-143)*. Boston: Houghton Miffin.

Gili Fivela, B. (2012). Testing the perception of L2 intonation. *Methodological Perspectives on Second Language Prosody. Papers from ML2P 2012*, 17–30.

Gili Fivela, Barbara, Avesani, C., Barone, M., Bocci, G., Crocco, C., D'Imperio, M., … Sorianello, P. (2015). Intonational phonology of the regional varieties of Italian. In *Intonation in Romance* (pp. 140–197). Oxford University Press.

Girju, R., Badulescu, A., & Moldovan, D. (2006). Automatic discovery of part-whole relations. *Computational Linguistics*, *32*(1), 83–135.

Goldsmith, J. (1976). An overview of autosegmental phonology. *Linguistic Analysis*, (2), 23–68.

Golestani, N., & Zatorre, R. J. (2009). Individual differences in the acquisition of second language phonology. *Brain and Language*, *109*(2), 55–67.

Goudbeek, M., Smits, R., Cutler, A., & Swingley, D. (2005). Acquiring auditory and phonetic categories. In *Handbook of categorization in cognitive science* (pp. 497–513). Elsevier.

Grabe, E. (1998). Pitch accent realization in English and German. *Journal of Phonetics*, *26*(2), 129–143.

Grabe, E., Kochanski, G., & Coleman, J. (2005). The intonation of native accent varieties in the British Isles-potential for miscommunication. English pronunciation models: A changing scene, 21. In *English pronunciation models: A changing scene*.

Grabe, E., Post, B., Nolan, F., & Farrar, K. (2000). Pitch accent realization in four varieties of British English. *Journal of Phonetics*, *28*(2), 161–185.

Grabe, E., Rosner, B. S., Garcia-Albea, J. E., & Zhou, X. (2003). Perception of English intonation by English, Spanish, and Chinese listeners. *Language and Speech*, (46(4)), 375–401.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.

Green, H., & Tobin, Y. (2009). Prosodic analysis is difficult… but worth it: A study in high functioning autism. *International Journal of Speech-Language Pathology*, *11*(4), 308–315.

Greenberg, J. H. (1966). *Universals of Language. With special attention to feature hierarchies*. La Haya, Holanda, Mouton.

Grice, M. (2017). *The intonation of interrogation in Palermo Italian: implications for intonation theory* (Vol. 334). Walter de Gruyter GmbH & Co KG.

Grice, M., & Baumann, S. (2007). An introduction to intonation-functions and models. *TRENDS IN LINGUISTICS STUDIES AND MONOGRAPHS*, *186*, 25.

Grice, M., Baumann, S., & Benzmüller, R. (2006). Autosegmental-Metrical Phonology. *Prosodic Typology: The Phonology of Intonation and Phrasing*, *1*, 55.

Grice, M., D'Imperio, M., Savino, M., & Avesani, C. (2005). Strategies for intonation labelling across varieties of Italian. *Prosodic Typology: The Phonology of Intonation and Phrasing*, 362–389.

Grice, M., Reyelt, M., Benzmuller, R., Mayer, J., & Batliner, A. (1996). Consistency in transcription and labelling of German intonation with GToBI. In *Spoken Language, 1996. ICSLP 96. Proceedings., Fourth International Conference on* (Vol. 3, pp. 1716–1719). IEEE.

Grice, M., & Savino, M. (1995). Intonation and communicative function in a regional variety of Italian. *Phonus*, *1*, 19–32.

Grønnum, N. (1994). Rhythm, duration and pitch in regional variants of standard Danish. *Acta Linguistica Hafniensia*, *27*(1), 189–218.

Guiora, A. Z. (1967). Toward a systematic study of empathy. *Comprehensive Psychiatry*, *8*(5), 375–385.

Guiora, A. Z., Brannon, R. C., & Dull, C. Y. (1972). Empathy And Second Language Learning. *Language Learning*, *22*(1), 111–130.

Guiora, A. Z., Lane, H., & Bosworth, L. A. (1967). An exploration of some personality variables in authentic pronunciation of a second language. *Studies in Language and Language Behavior*, *4*, 261–266.

Gundel, J. K., & Fretheim, T. (2004). Topic and focus. *The Handbook of Pragmatics*, *175*, 196.

Gundel, J. K., Hedberg, N., & Zacharski, R. (1993). Cognitive status and the form of referring expressions in discourse. *Language*, 274–307.

Gundel, J. K., Houlihan, K., & Sanders, G. A. (1986). Markedness and distribution in phonology and syntax. In *Markedness* (pp. 107–138). Springer.

Gussenhoven, C. (1983). Focus, mode and the nucleus. *Journal of Linguistics*, *19*(2), 377–417.

Gussenhoven, C. (1984). *On the grammar and semantics of sentence accents*. Dordrecht: Foris.

Gussenhoven, C. (1999). Discreteness and gradience in intonational contrasts. *Language and Speech*, *42*(2–3), 283–305.

Gussenhoven, C. (2002). Intonation and interpretation: phonetics and phonology. In *Speech Prosody 2002, International Conference*.

Gussenhoven, C. (2006). Experimental approaches to establishing discreteness of intonational contrasts. *Methods in Empirical Prosody Research*, 321–334.

Gut, U. (2007). Learner corpora in second language prosody research and teaching. *TRENDS IN LINGUISTICS STUDIES AND MONOGRAPHS*, *186*, 145.

Hajicová, E. (1980). A Dependency Based Specification of Topic and Focus-Background and Motivation. *SMIL: Journal of Linguistic Calculus Stockholm*, (1–2), 93–109.

Hajičová, E., Sgall, P., & Skoumalová, H. (1995). An automatic procedure for topic-focus identification. *Computational Linguistics*, *21*(1), 81–94.

Hall, T. A. (1999). The phonological word: a review. *Studies on the Phonological Word*, *174*, 1–22.

Halliday, M. A. (1967). Notes on transitivity and theme in English: Part 2. *Journal of Linguistics*, *3*(2), 199–244.

Halliday, M. A. K. (1967). *Intonation and grammar in British English* (Vol. 48). Walter de Gruyter GmbH & Co KG.

Hamilton, A. F. de C. (2009). Research review: Goals, intentions and mental states: Challenges for theories of autism. *Journal of Child Psychology and Psychiatry*, *50*(8), 881–892.

Hanson, C., & Hanson, S. J. (2005). Categorization in neuroscience: brain response to objects and events. In *Handbook of categorization in cognitive science* (pp. 119–140). Elsevier.

Happé, F. (1995). Understanding minds and metaphors: Insights from the study of figurative language in autism. *Metaphor and Symbol*, *4*, 275–295.

Harding, S., Cooke, M., & König, P. (2007). Auditory gist perception: an alternative to attentional selection of auditory streams? In *International Workshop on Attention in Cognitive Systems* (pp. 399–416). Springer.

Harley, B. (1986). *Age in second language acquisition* (Vol. 22). College-Hill.

Harley, B., & Hart, D. (1997). Language aptitude and second language proficiency in classroom learners of different starting ages. *Studies in Second Language Acquisition*, 379–400.

Harnad, S. (2003). Categorical perception.

Harnad, S., Cohen, H., & Lefebvre, C. (2005). *Handbook of categorization in cognitive science*. Oxford: Elsevier, chapter Cognition is categorization.

Hayes, B. P. (1980). *A metrical theory of stress rules*. Massachusetts Institute of Technology.

Heald, S. L., Van Hedger, S. C., & Nusbaum, H. C. (2017). Perceptual plasticity for auditory object recognition. *Frontiers in Psychology*, *8*, 781.

Heeger, D. (1998). *Signal detection theory*. California.

Heim, S., & Alter, K. (2006). Prosodic pitch accents in language comprehension and production: ERP data and acoustic analyses. *Acta Neurobiologiae Experimentalis*, *66*(1), 55.

Hickok, G., & Poeppel, D. (2000). Towards a functional neuroanatomy of speech perception. *Trends in Cognitive Sciences*, *4*(4), 131–138.

Hickok, G., & Poeppel, D. (2004). Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition*, *92*(1), 67–99.

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews. Neuroscience*, *8*(5), 393.

Hirschberg, J., & Ward, G. (1992). The influence of pitch range, duration, amplitude and spectral features on the interpretation of the rise-fall-rise intonation contour in English. *Journal of Phonetics*.

Hirst, D., & Di Cristo, A. (1998). *Intonation systems: a survey of twenty languages*. Cambridge University Press.

Holt, L. L., & Lotto, A. J. (2010). Speech perception as categorization. *Attention, Perception, & Psychophysics*, *72*(5), 1218–1227.

Holyoak, K. J., & Thagard, P. (1995). *Mental leaps*.

House, D. (1990). *Tonal perception in speech* (Vol. 24). Lund University Press.

House, D. (1996). Differential perception of tonal contours through the syllable. In *Spoken Language, 1996. ICSLP 96. Proceedings., Fourth International Conference on* (Vol. 4, pp. 2048–2051). IEEE.

House, D. (1999). Perception of pitch and tonal timing: implications for mechanisms of tonogenesis. In *Proceedings of the 14th International Congress of Phonetic Sciences* (pp. 1823–1826).

Hruska, C., & Alter, K. (2004). Prosody in dialogues and single sentences: How prosody can influence speech perception. In *Information structure: Theoretical and empirical aspects* (pp. 221–226).

Hruska, C., Alter, K., Steinhauer, K., & Steube, A. (2001). Misleading dialogues: Human's brain reaction to prosodic information. *Orality and Gesture. Interactions et Comportements Multimodaux Dans La Communication. Paris. L'Harmattan*, 425–430.

Hyltenstam, K., & Abrahamsson, N. (2000). Who can become native-like in a second language? All, some, or none? On the maturational constraints controversy in second language acquisition. *Studia Linguistica*, *54*(2), 150–166.

Hyman, L. M. (1975). *Phonology: theory and analysis*. Harcourt College Pub.

Ito, K., & Garnsey, S. M. (2004). Brain responses to focus-related prosodic mismatch in Japanese. In *Speech Prosody 2004, International Conference*.

Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *20*(11), 1254–1259.

Jackendoff, R. S. (1972). Semantic interpretation in generative grammar.

Jacobs, J. (1984). Funktionale Satzperspektive und Illokutionssemantik. *Linguistische Berichte*, pp. 25–58.

Jakobson, R. (1941). *Child Language, Aphasia, and Phonological Universals* (Mouton). The Hague.

Jenner, B. R. (1976). Interlanguage and foreign accent. *Interlanguage Studies Bulletin*, 166–195.

Jilka, M. (2000). *The contribution of intonation to the perception of foreign accent*. Universitat Stuttgart.

Johnson, J. S. (1992). Critical period effects in second language acquisition: The effect of written versus auditory materials on the assessment of grammatical competence. *Language Learning*, *42*(2), 217–248.

Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology*, *21*(1), 60–99.

Johnson, J. S., & Newport, E. L. (1991). Critical period effects on universal properties of language: The status of subjacency in the acquisition of a second language. *Cognition*, *39*(3), 215–258.

Johnson, K. (1997). Speech perception without speaker normalization: An exemplar model. *Talker Variability in Speech Processing*, 145–165.

Jun, S.-A., & Oh, M. (2000). Acquisition of second language intonation. In *INTERSPEECH* (pp. 73–76).

Jusczyk, P. W. (1997). Finding and remembering words: Some beginnings by English-learning infants. *Current Directions in Psychological Science*, *6*(6), 170–174.

Kalinli, O., & Narayanan, S. S. (2007). A saliency-based auditory attention model with applications to unsupervised prominent syllable detection in speech. In *Eighth Annual Conference of the International Speech Communication Association*.

Kingdom, F. A., & Prins, N. (2010). *Psychophysics: A practical introduction*. Academic Press London.

Kluender, K. R., Lotto, A. J., Holt, L. L., Greenberg, S., & Ainsworth, W. (2005). Listening to speech: An auditory perspective.

Kohler, K. J. (1987). Categorical pitch perception. In *Proceedings of the XIth International Congress of Phonetic Sciences* (Vol. 5, pp. 331–333).

Kohler, K. J. (1991). *Studies in German intonation*. Institut für Phonetik und digitale Sprachverarbeitung, Universität Kiel.

Kohler, K. J. (1997). Modelling prosody in spontaneous speech. In *Computing Prosody* (pp. 187–210). Springer.

Kornai, A. (1998). Analytic models in phonology. In *The Organization of Phonology: Constraints, Levels and Representations* (Oxford University Press, pp. 395–418).

Krashen, S. (1994). The input hypothesis and its rivals. In *Implicit and explicit learning of languages* (pp. 45–77).

Krashen, S. (1999). Seeking a role for grammar. A review of some recent studies (Vol. 2, pp. 245–254).

Krashen, S. D., Long, M. A., & Scarcella, R. C. (1979). Age, rate and eventual attainment in second language acquisition. *TESOL Quarterly*, 573–582.

Krumhansl, C. L. (1991). Music psychology: Tonal structures in perception and memory. *Annual Review of Psychology*, *42*(1), 277–303.

Kuhl, P. K., & Miller, J. D. (1975). Speech perception by the chinchilla: voice-voiceless distinction in alveolar plosive consonants. *Science*, *190*, 69–72.

Kuhl, P. K., & Miller, J. D. (1978). Speech perception by the chinchilla: identification functions for synthetic VOT stimuli. *J. Acoust. Soc. Am.*, *63*, 905–917. https://doi.org/10.1121/1.381770

Kuhl, Patricia K. (1991). Human adults and human infants show a “perceptual magnet effect” for the prototypes of speech categories, monkeys do not. *Attention, Perception, & Psychophysics*, *50*(2), 93–107.

Kuhl, Patricia K. (2000). A new view of language acquisition. *Proceedings of the National Academy of Sciences*, *97*(22), 11850–11857.

Kuno, S. (1978). Generative discourse analysis in America. *Current Trends in Textlinguistics*, 275–294.

Lacerda, F. (1995). The perceptual magnet effect: an emergent consequence of exemplar-based phonetic memory. *Proc. Int. Congr. Phonetic Sci.*, *2*, 140–147.

Ladd, D. R. (2008). *Intonational phonology*. Cambridge University Press.

Ladd, R. D. (1980). *The structure of intonational meaning: evidence from English. Bloomington: Indiana University Press.* Bloomington: Indiana University Press.

Ladd, R. D. (1986). Intonational phrasing: the case for recursive prosodic structure. *Phonology*, *3*, 311–340.

Ladd, R. D. (1996). *Intonational phonology*. Cambridge University Press.

LaGasse, L. L., Neal, A. R., & Lester, B. M. (2005). Assessment of infant cry: acoustic cry analysis and parental perception. *Developmental Disabilities Research Reviews*, *11*(1), 83–93.

Lalumera, E. (2005). Conceptual Analysis and Philosophical Naturalism. In *Handbook of Categorization in Cognitive Science* (pp. 1055–1071). Elsevier.

Lambrecht, K. (1996). *Information structure and sentence form: Topic, focus, and the mental representations of discourse referents* (Vol. 71). Cambridge University Press.

Landgraf, K. (2003). *Steigende Intonationskonturenim Deutschen: Experimentalphonetische Untersuchungen zur auditiven Kategorisierung.* Universität zu Kiel.

Leekam, S. R., Hunnisett, E., & Moore, C. (1998). Targets and cues: Gaze-following in children with autism. *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, *39*(7), 951–962.

Leekam, S. R., López, B., & Moore, C. (2000). Attention and joint attention in preschool children with autism. *Developmental Psychology*, *36*(2), 261.

Lenneberg, E. H. (1967). The biological foundations of language. *Hospital Practice*, *2*(12), 59–67.

Lepetit, D. (1989). Cross-linguistic influence in intonation: French/Japanese and French/English. *Language Learning*, *39*(3), 397–413.

Lepschy, A. L., & Lepschy, G. (2013). *The Italian language today*. Routledge.

Leslie, A. M. (1991). The theory of mind impairment in autism: Evidence for a modular mechanism of development? In *Natural theories of mind: Evolution, development and simulation of everyday mindreading* (pp. 63–78). Cambridge, MA, US: Basil Blackwell.

Liang, J., & van Heuven, V. J. (2007). Chinese tone and intonation perceived by L1 and L2 listeners. *Tones and Tunes*, *2*, 27–61.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychol. Rev.*, *74*, 431–461.

Liberman, Alvin M. (1957). Some results of research on speech perception. *The Journal of the Acoustical Society of America*, *29*(1), 117–123.

Liberman, Alvin M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, *74*(6), 431.

Liberman, Alvin M., Harris, K. S., Kinney, J. A., & Lane, H. (1961). The discrimination of relative onset-time of the components of certain speech and nonspeech patterns. *Journal of Experimental Psychology*, *61*(5), 379.

Liberman, Alvin M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, *21*(1), 1–36.

Liberman, Alvin Meyer. (1996). *Speech: A special code*. MIT press.

Liberman, M., & Prince, A. (1977). On stress and linguistic rhythm. *Linguistic Inquiry*, *8*(2), 249–336.

Liberman, M., & Sag, I. (1974). Prosodic form and discourse function. In *Chicago Linguistics Society* (Vol. 10, pp. 416–427).

Liberman, M. Y. (1975). *The intonational system of English.* Massachusetts Institute of Technology.

Ling, L. E., & Grabe, E. (1999). A contrastive study of prosody and lexical stress placement in Singapore English and British English. *Language and Speech*, *42*(1), 39–56.

Livingston, K. R., Andrews, J. K., & Harnad, S. (1998). Categorical perception effects induced by category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(3), 732.

Long, M. H. (1990). Maturational constraints on language development. *Studies in Second Language Acquisition*, *12*(3), 251–285.

Long, M. H. (2008). 16 Stabilization and Fossilization in Interlanguage Development. *The Handbook of Second Language Acquisition*, *27*, 487.

Lotto, A. J., & Holt, L. L. (2016). Speech Perception: The View from the Auditory System. In *Neurobiology of Language* (pp. 185–194). Elsevier.

Macmillan, N. A., & Creelman, C. D. (1991). Detection theory: A user's guideCambridge University Press. *New York*.

Magne, C., Astésano, C., Lacheret-Dujour, A., Morel, M., Alter, K., & Besson, M. (2005). On-line processing of "pop-out" words in spoken French dialogues. *Journal of Cognitive Neuroscience*, *17*(5), 740–756.

Maiden, M., & Robustelli, C. (2000). A Reference Grammar of Modern Italian. *London: Arnold*.

Martin-Rhee, M., & Bialystok, E. (2008). The development of two types of inhibitory control in monolingual and bilingual children. *Bilingualism: Language and Cognition*, *11*(1), 81–93.

Mathesius, V. (1936). On some problems of the systematic analysis of grammar. *Travaux Du Cercle Linguistique de Prague*, *6*(96), 107.

Mattingly, I. G., Liberman, A. M., Syrdal, A. K., & Halwes, T. (1971). Discrimination in speech and nonspeech modes. *Cognitive Psychology*, *2*(2), 131–157.

Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, *82*(3), B101–B111.

McAllister, R., Flege, J. E., & Piske, T. (2002). The influence of L1 on the acquisition of Swedish quantity by native speakers of Spanish, English and Estonian. *Journal of Phonetics*, *30*(2), 229–258.

Mennen, I. (1999). The realisation of nucleus placement in second language intonation. In *Proceedings of the fourteenth international congress of phonetic sciences* (pp. 555–558).

Mennen, I. (2004). Bi-directional interference in the intonation of Dutch speakers of Greek. *Journal of Phonetics*, *32*(4), 543–563.

Mennen, I. (2007). Phonological and phonetic influences in non-native intonation. *TRENDS IN LINGUISTICS STUDIES AND MONOGRAPHS*.

Mennen, I. (2015). Beyond segments: Towards a L2 intonation learning theory. In *Prosody and language in contact* (pp. 171–188). Springer.

Mennen, I., Schaeffler, F., & Docherty, G. (2012). Cross-language differences in fundamental frequency range: A comparison of English and German. *The Journal of the Acoustical Society of America*, *131*(3), 2249–2260.

Meyer, M., Steinhauer, K., Alter, K., Friederici, A. D., & von Cramon, D. Y. (2004). Brain activity varies with modulation of dynamic pitch variance in sentence melody. *Brain and Language*, *89*(2), 277–289.

Miller, J. D., Wier, C. C., Pastore, R. E., Kelly, W. J., & Dooling, R. J. (1976). Discrimination and labeling of noise–buzz sequences with varying noise-lead times: An example of categorical perception. *The Journal of the Acoustical Society of America*, *60*(2), 410–417.

Mirman, D., Holt, L. L., & McClelland, J. L. (2004). Categorization and discrimination of nonspeech sounds: Differences between steady-state and rapidly-changing acoustic cues. *The Journal of the Acoustical Society of America*, *116*(2), 1198–1207.

Miyake, A., & Shah, P. (1999). *Models of working memory: Mechanisms of active maintenance and executive control* (Cambridge University Press.).

Morris, R. K., & Folk, J. R. (1998). Focus as a contextual priming mechanism in reading. *Memory & Cognition*, *26*(6), 1313–1322.

Moulton, W. G. (1962). Toward a classification of pronunciation errors. *The Modern Language Journal*, *46*(3), 101–109.

Munro, M. J., & Derwing, T. M. (1995). Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. *Language and Speech*, *38*(3), 289–306.

Munro, M. J., Flege, J. E., & MacKay, I. R. (1996). The effects of age of second language learning on the production of English vowels. *Applied Psycholinguistics*, *17*(3), 313–334.

Muysken, P. (2005). A modular approach to grammatical categories evidence from language diversity and contact. In *Handbook of Categorization in Cognitive Science* (pp. 45–69). Elsevier.

Näätänen, R. (1990). The role of attention in auditory information processing as revealed by event-related potentials and other brain measures of cognitive function. *Behavioral and Brain Sciences*, *13*(2), 201–233.

Näätänen, R., Kujala, T., & Winkler, I. (2011). Auditory processing that leads to conscious perception: a unique window to central auditory processing opened by the mismatch negativity and related responses. *Psychophysiology*, *48*(1), 4–22.

Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: a review. *Clinical Neurophysiology*, *118*(12), 2544–2590.

Nadig, A., & Shaw, H. (2012). Acoustic and perceptual measurement of expressive prosody in high-functioning autism: Increased pitch range and what it means to listeners. *Journal of Autism and Developmental Disorders*, *42*(4), 499–511.

Nauchi, A., & Sakai, K. L. (2009). Greater leftward lateralization of the inferior frontal gyrus in second language learners with higher syntactic abilities. *Human Brain Mapping*, *30*(1), 3625–3635.

Nespor, M., & Vogel, I. (1986). Prosodic phonology (studies in generative grammar 28). *Dordrecht: Foris*.

Nibert, Holly J. (2005). The acquisition of the phrase accent by intermediate and advanced adult learners of Spanish as a second language. In *Selected Proceedings of the 6th Conference on the Acquisition of Spanish and Portuguese as First and Second Languages* (pp. 108–122). Citeseer.

Nibert, Holly J. (2006). The acquisition of the phrase accent by beginning adult learners of Spanish as a second language. In *Selected proceedings of the 2nd conference on laboratory approaches to Spanish phonetics and phonology* (pp. 131–148).

Nibert, Holly Joy. (2000). *Phonetic and phonological evidence for intermediate phrasing in Spanish intonation*. University of Illinois at Urbana-Champaign.

Niebuhr, O. (2003). Perceptual study of timing variables in F0 peaks.

Niebuhr, O. (2006). The role of the accented-vowel onset in the perception of German early and medial peaks. In *Speech Prosody* (pp. 109–112).

Niebuhr, O. (2007). Categorical perception in intonation: a matter of signal dynamics?

Niebuhr, O., & Kohler, K. J. (2004). Perception and cognitive processing of tonal alignment in German. In *International Symposium on Tonal Aspects of Languages: With Emphasis on Tone Languages*.

Obler, L. K., & Fein, D. E. (1988). *The exceptional brain: Neuropsychology of talent and special abilities* (Guillford Press).

O'Connor, J. D., & Arnold, G. F. (1973). *Intonation of Colloquial English* (2nd.). London: Longmans.

Ohala, J. J. (1994). The frequency code underlines the sound symbolic use of voice of pitch. In *Sound symbolism*. Cambridge: Cambridge University Press.

Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science*, *308*(5719), 255–258.

Oyama, S. (1976). A sensitive period for the acquisition of a nonnative phonological system. *Journal of Psycholinguistic Research*, *5*(3), 261–283.

Palmeri, T. J., & Gauthier, I. (2004). Visual object understanding. *Nature Reviews Neuroscience*, *5*(4), 291.

Paradis, M. (1994). Neurolinguistic aspects of implicit and explicit memory: implication for bilingualism and second language acquisition. In *Implicit and explicit learning of languages.*

Paradis, M. (2009). *Declarative and procedural determinants of second languages John Benjamins Publishing* (John Benjamins Publishing, Vol. 40).

Partee, B. (1998). Copula inversion puzzles in English and Russian. In *Formal Approaches to Slavic Linguistics: The Seattle Meeting* (pp. 361–395).

Patel, A. D. (2010). *Music, language, and the brain*. Oxford University Press.

Patel, A. D., Foxton, J. M., & Griffiths, T. D. (2005). Musically tone-deaf individuals have difficulty discriminating intonation contours extracted from speech. *Brain and Cognition*, *59*(3), 310–313.

Patkowski, M. S. (1980). The sensitive period for the acquisition of syntax in a second language 1. *Language Learning*, *30*(2), 449–468.

Patterson, D. (2000). *A Linguistic Approach to Pitch Range Modelling. Ph.D. Dissertation*. University of Edinburgh.

Peretz, I. (2001). Brain specialization for music (Vol. 1, pp. 153–165).

Peretz, Isabelle, & Hyde, K. I. (2003). What is specific to music processing? Insights from congenital amusia. *Trends in Cognitive Sciences*, *7*(8), 362–367.

Pierrehumbert, J. (2002). Word-specific phonetics. *Laboratory Phonology*, *7*.

Pierrehumbert, J. (2003). Phonetic diversity, statistical learning, and acquisition of phonology. *Language and Speech*, *46*(2–3), 115–154.

Pierrehumbert, Janet B., & Steele, S. A. (1989). Categories of tonal alignment in English. *Phonetica*, (46(4)), 181–196.

Pierrehumbert, Janet Breckenridge. (1980). *The phonology and phonetics of English intonation*. Massachusetts Institute of Technology.

Pierrehumbert, Janet, & Hirschberg, J. B. (1990). The meaning of intonational contours in the interpretation of discourse. *Intentions in Communication*, 271–311.

Pisoni, D. B. (1977). Identification and discrimination of the relative onset time of two component tones: implications for voicing perception in stops. *J. Acoust. Soc. Am.*, *61*, 1352–1361. https://doi.org/10.1121/1.381409

Pisoni, David B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Attention, Perception, & Psychophysics*, *13*(2), 253–260.

Pisoni, David B. (1975). Auditory short-term memory and vowel perception. *Memory & Cognition*, *3*(1), 7–18.

Poeppel, D., & Hickok, G. (2004). *Towards a new functional anatomy of language*. Elsevier.

Polka, L., & Werker, J. F. (1994). Developmental changes in perception of nonnative vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance*, *20*(2), 421.

Polster, M. R., & Rose, S. B. (1998). Disorders of auditory processing: evidence for modularity in audition. *Cortex*, *34*(1), 47–65.

Post, B. M. B., D'Imperio, M., & Gussenhoven, C. (2007). Fine phonetic detail and intonational meaning.

Prince, E. F. (1981). Toward a taxonomy of given-new information. *Radical Pragmatics*.

Rasier, L., Caspers, J., & van Heuven, V. (2010). Accentual marking of information status in Dutch and French as foreign languages. Production and perception.

Rasier, L., & Hiligsmann, P. (2007). Prosodic transfer from L1 to L2. Theoretical and methodological issues. *Nouveaux Cahiers de Linguistique Française*, *28*(2007), 41–66.

Reber, A. S. (1993). *Implicit Learning and Tacit Knowledge: An Essay on the Cognitive Unconscious (Oxford Psychology Series*.

Redi, L. (2003). Categorical effects in production of pitch contours in English. In *Proceedings of the 15th International Congress of the Phonetic Sciences* (pp. 2921–2924).

Reinhart, T. (1982). Pragmatics and linguistics an analysis of sentence topics.

Reiterer, S. M., Hu, X., Erb, M., Rota, G., Nardo, D., Grodd, W., … Ackermann, H. (2011). Individual differences in audio-vocal speech imitation aptitude in late bilinguals: functional neuro-imaging and brain morphology. *Frontiers in Psychology*, *2*.

Remijsen, B., & van Heuven, V. J. (1999). Gradient and categorical pitch dimensions in Dutch: diagnostic test. In *Proceedings of the 14th International Congress of Phonetic Sciences* (Vol. 2, pp. 1865–1868).

Repp, B. H. (1984). Categorical perception: Issues, methods, findings. In *Speech and language* (Vol. 10, pp. 243–335). Elsevier.

Rey, G. (2005). Philosophical analysis as cognitive psychology: The case of empty concepts. In *Handbook of categorization in cognitive science* (pp. 71–89). Elsevier.

Richards, J. (1974). *Error analysis: perspectives on second language learning*. London: Longman.

Ringbom, H. (2007). *Cross-linguistic similarity in foreign language learning* (Vol. 21). Multilingual Matters.

Robert, S. (2005). Categorization, reasoning, and memory from a neo-logical point of view. In *Handbook of categorization in cognitive science* (pp. 699–717). Elsevier.

Robinson, P. (1995). Attention, memory, and the "noticing" hypothesis. *Language Learning*, *45*(2), 283–331.

Robinson, P., Mackey, A., Gass, S. M., & Schmidt, R. (2012). Attention and awareness in second language acquisition. In *The Routledge handbook of second language acquisition* (pp. 247–267).

Roehr, C. T., & Baumann, S. (2010). Prosodic marking of information status in German. In *Speech Prosody 2010-Fifth International Conference*.

Röhr, C., & Baumann, S. (2011). Decoding information status by type and position of accent in German. In *Proceedings 17th ICPhS, Hong Kong, China* (pp. 1706–1709).

Rota, G., & Reiterer, S. (2009). Cognitive aspects of pronunciation talent. In *Language talent and brain activity* (Vol. 1, pp. 67–96).

Ruta, L., Mazzone, D., Mazzone, L., Wheelwright, S., & Baron-Cohen, S. (2012). The Autism-Spectrum Quotient—Italian version: A cross-cultural confirmation of the broader autism phenotype. *Journal of Autism and Developmental Disorders*, *42*(4), 625–633.

Rutherford, M. D., Baron-Cohen, S., & Wheelwright, S. (2002). Reading the mind in the voice: A study with normal adults and adults with Asperger syndrome and high functioning autism. *Journal of Autism and Developmental Disorders*, *32*(3), 189–194.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month old infants. *Science*, *274*, 1926–1928. https://doi.org/10.1126/science.274.5294.1926

Sag, I., & Liberman, M. (1975). The intonational disambiguation of indirect speech acts. In *Proceedings of the Chicago Linguistics Society* (Vol. 11, pp. 487–497).

Sanford, A. J., & Garrod, S. C. (1981). *Understanding written language: Explorations of comprehension beyond the sentence*. John Wiley & Sons.

Savino, M. (2009). Intonational features for identifying regional accents of Italian. In *Tenth Annual Conference of the International Speech Communication Association*.

Savino, M. (2012). The intonation of polar questions in Italian: Where is the rise? *Journal of the International Phonetic Association*, *42*(1), 23–48.

Schachter, J. (1990). On the issue of completeness in second language acquisition. *Interlanguage Studies Bulletin (Utrecht)*, *6*(2), 93–124.

Schachter, J. (1996). Learning and triggering in adult L2 acquisition. *Performance and Competence in Second Language Acquisition*, 70–88.

Schacter, J., Rounds, P., Wright, S., & Smith, T. (1996). Comparing conditions for learning syntactic patterns: Attentional, non attentional and aware. *Technical Report*, 96–108.

Schepman, A., Lickley, R., & Ladd, D. R. (2006). Effects of vowel length and “right context” on the alignment of Dutch nuclear accents. *Journal of Phonetics*, *34*(1), 1–28.

Schmidt, R. (1992). Awareness and second language acquisition. Annual review of applied linguistics (pp. 206–226).

Schneider, K. (2012). The German boundary tones: categorical Perception, perceptual magnets, and the perceptual reference space.

Schneider, K., Dogil, G., & Möbius, B. (2009). German boundary tones show categorical perception and a perceptual magnet effect when presented in different contexts. In *Tenth Annual Conference of the International Speech Communication Association*.

Schneider, K., & Lintfert, B. (2003). Categorical perception of boundary tones in German. In *Proceedings of the 15th International Conference of the Phonetic Sciences* (pp. 631–634).

Schneider, K., & Möbius, B. (2005). Perceptual magnet effect in German boundary tones. In *Ninth European Conference on Speech Communication and Technology*.

Schumacher, P. B., & Baumann, S. (2010). Pitch accent type affects the N400 during referential processing: *NeuroReport*, *21*(9), 618–622.

Schwartz, B. D., & Sprouse, R. (1994). Word order and nominative case in nonnative language acquisition: a longitudinal study of (L1 Turkish) German interlanguage. In *Language acquisition studies in generative grammar,* (pp. 71–89).

Scovel, T. (2000). A critical review of the critical period research. *Annual Review of Applied Linguistics*, *20*, 213–223.

Seliger, H. W. (1978). Implications of a multiple critical periods hypothesis for second language learning. *Second Language Acquisition Research: Issues and Implications*, 11–19.

Selinker, L. (1972). Interlanguage. *IRAL-International Review of Applied Linguistics in Language Teaching*, *10*(1–4), 209–232.

Selinker, L., & Lamendella, J. T. (1978). Two perspectives on fossilization in interlanguage learning. *Interlanguage Studies Bulletin*, 143–191.

Selkirk, E. (1980). *On prosodic structure and its relation to syntactic structure*. Indiana University Linguistics Club.

Selkirk, E. (1984). On the major class features and syllable theory. *Language Sound Structure*.

Selkirk, E. (1986). On derived domains in sentence phonology. *Phonology*, *3*(1), 371–405.

Sgall, P., Hajicová, E., & Panevova, J. (1986). *The meaning of the sentence in its semantic and pragmatic aspects*. Springer Science & Business Media.

Shackleton, T. M., & Carlyon, R. P. (1994). The role of resolved and unresolved harmonics in pitch perception and frequency modulation discrimination. *The Journal of the Acoustical Society of America*, *95*(6), 3529–3540.

Sharda, M., Subhadra, T. P., Sahay, S., Nagaraja, C., Singh, L., Mishra, R., … Singh, N. C. (2010). Sounds of melody—Pitch patterns of speech in autism. *Neuroscience Letters*, *478*(1), 42–45.

Sheinkopf, S. J., Iverson, J. M., Rinaldi, M. L., & Lester, B. M. (2012). Atypical Cry Acoustics in 6-Month-Old Infants at Risk for Autism Spectrum Disorder. *Autism Research*, *5*(5), 331–339.

Shriberg, L. D., Paul, R., McSweeny, J. L., Klin, A., Cohen, D. J., & Volkmar, F. R. (2001). Speech and prosody characteristics of adolescents and adults with high-functioning autism and Asperger syndrome. *Journal of Speech, Language, and Hearing Research*, *44*(5), 1097–1115.

Siegal, M., & Peterson, C. C. (2008). Language and theory of mind in atypically developing children: evidence from studies of deafness, blindness and autism. In *Social Cognition And Developmental Psychopathology* (pp. 81–114). United States: Sharp, CarlaFonagy, PeterGoodyer, Ian.

Simons, P. (1987). *Parts: A Study in Ontology*. New York: Oxford University Press.

Simons, P. (1991). Part/whole II: Mereology since 1900. In *Handbook of Metaphysics and Ontology* (pp. 672–675). Munich: Philosophia.

Sirviö, P., & Michelsson, K. (1976). Sound-spectrographic cry analysis of normal and abnormal newborn infants. *Folia Phoniatrica et Logopaedica*, *28*(3), 161–173.

Skehan, P. (2002). Theorising and updating aptitude. Individual differences and instructed language learning. In *Individual differences and instructed language learning* (Vol. 2, pp. 69–94).

Skehan, P. (2013). Language aptitude. In *The Routledge handbook of second language acquisition* (Routledge, pp. 381–395).

Snow, C. E., & Hoefnagel-Höhle, M. (1978). The critical period for language acquisition: Evidence from second language learning. *Child Development*, 1114–1128.

Sorace, A. (1993). Incomplete vs. divergent representations of unaccusativity in non native grammars of Italian. *Second Language Research*, *9*(1), 22–47.

Sowa, J. F. (2005). Categorization in cognitive computer science. In *Handbook of categorization in cognitive science* (pp. 141–163). Elsevier.

Stagray, J. R., Downs, D., & Sommers, R. K. (1992). Contributions of the fundamental, resolved harmonics, and unresolved harmonics in tone-phoneme identification. *Journal of Speech, Language, and Hearing Research*, *35*(6), 1406–1409.

Steinhauer, K. (2003a). Electrophysiological correlates of prosody and punctuation. *Brain and Language*, *86*(1), 142–164.

Steinhauer, K. (2003b). Electrophysiological correlates of prosody and punctuation. *Brain and Language*, *86*(1), 142–164.

Steinhauer, K., Drury, J. E., Portner, P., Walenski, M., & Ullman, M. T. (2010). Syntax, concepts, and logic in the temporal dynamics of language comprehension: Evidence from event-related potentials. *Neuropsychologia*, *48*(6), 1525–1542.

Steinhauer, K., & Friederici, A. D. (2001). Prosodic boundaries, comma rules, and brain responses: The closure positive shift in ERPs as a universal marker for prosodic phrasing in listeners and readers. *Journal of Psycholinguistic Research*, *30*(3), 267–295.

Stolterfoht, B., Friederici, A. D., Alter, K., & Steube, A. (2007). Processing focus structure and implicit prosody during reading: Differential ERP effects. *Cognition*, *104*(3), 565–590.

Strange, W. (2007). Cross-language phonetic similarity of vowels. *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege*, *17*.

Studdert-Kennedy, M., Liberman, A. M., Harris, K. S., & Cooper, F. S. (1970). Motor theory of speech perception: A reply to Lane's critical review.

Sturt, P., Sanford, A. J., Stewart, A., & Dawydiak, E. (2004). Linguistic focus and good-enough representations: An application of the change-detection paradigm. *Psychonomic Bulletin & Review*, *11*(5), 882–888.

Surian, L. (1996). Are children with autism deaf to Gricean maxims? *Cognitive Neuropsychiatry*, *1*(1), 55–72.

Surian, L., & Siegal, M. (2008). Language and communication disorders in autism and Asperger syndrome. *Handbook of the Neuroscience of Language*, 377–384.

Swerts, M., Krahmer, E., & Avesani, C. (2002). Prosodic marking of information status in Dutch and Italian: A comparative analysis. *Journal of Phonetics*, *30*(4), 629–654.

Tahta, S., Wood, M., & Loewenthal, K. (1981). Foreign accents: Factors relating to transfer of accent from the first language to a second language. *Language and Speech*, *24*(3), 265–272.

Tannen, D. (1979). Processes and consequences of conversational style.

Taylor, S., & Wales, R. (1987). Primitive mechanisms of accent perception. *Journal of Phonetics*.

Toepel, U., Pannekamp, A., & Alter, K. (2007). Catching the news: Processing strategies in listening to dialogs as measured by ERPs. *Behavioral and Brain Functions*, *3*(1), 53.

Trim, J. L. (1988). Some contrastive intonated features of British English and German. *Essays on the English Language and Applied Linguistics on the Occasion of Gerhard Nickel's 60th Birthday., Heidelberg: Julius Groos*, 235–249.

Trimble, J. C. (2013). Perceiving intonational cues in a foreign language: Perception of sentence type in two dialects of Spanish. In *Selected proceedings of the 15th hispanic linguistics symposium* (pp. 78–92).

Trofimovich, P., & Baker, W. (2006). Learning second language suprasegmentals: Effect of L2 experience on prosody and fluency characteristics of L2 speech. *Studies in Second Language Acquisition*, *28*(1), 1–30.

Troubetzkoy, N. (1939). Phonemes and how to determine them. *Travaux Du Cercle Linguistique de Prague*.

Ueyama, M. (1997). The phonology and phonetics of second language intonation: the case of" Japanese English". In *Eurospeech*.

Uhmann, S. (1991). *Fokusphonologie.: Eine Analyse deutscher Intonationskonturen im Rahmen der nicht-linearen Phonologie.* (Vol. 252). Niemeyer Max Verlag GmbH.

Ullman, M. T. (2015). The declarative/procedural model: a neurobiological model of language learning, knowledge, and use. *Neurobiology of Language*, 953–968.

Vallduví, E. (1991). *The Informational Component: A Dissertation in Linguistics*. University of Pennsylvania, The Institute for Research in Cognitive Science.

van Boxtel, S., Bongaerts, T., & Coppen, P.-A. (2005). Native-like attainment of dummy subjects in Dutch and the role of the L1. *International Review of Applied Linguistics in Language Teaching*, *43*(4), 355–380.

van Heuven, V. J., & Kirsner, R. S. (2004). Phonetic or phonological contrasts in Dutch boundary tones? *Linguistics in the Netherlands*, *21*(1), 102–113.

Van Lacker, D. R., Cummings, J. L., Kreiman, J., & Dobkin, B. H. (1988). Phonagnosia: A dissociation between familiar and unfamiliar voices. *Cortex*, *24*(2), 195–209.

Von Békésy, G., & Wever, E. G. (1960). *Experiments in hearing* (Vol. 8). McGraw-Hill New York.

von Heusinger, K. (1999). *Intonation and Information Structure. The Representation of Focus in Phonology and Semantics. Habilitationssschrift, Universität Konstanz, Konstanz, Germany*.

Wales, R., & Taylor, S. (1987). Intonation cues to questions and statements: How are they perceived? *Language and Speech*, *30*(3), 199–211.

Wang, A. T., Lee, S. S., Sigman, M., & Dapretto, M. (2006). Neural basis of irony comprehension in children with autism: the role of prosody and context. *Brain*, *129*(4), 932–943.

Wang, L., Bastiaansen, M., Yang, Y., & Hagoort, P. (2011). The influence of information structure on the depth of semantic processing: How focus and pitch accent determine the size of the N400 effect. *Neuropsychologia*, *49*(5), 813–820.

Wang, L., Li, X., & Yang, Y. (2014). A review on the cognitive function of information structure during language comprehension. *Cognitive Neurodynamics*, *8*(5), 353–361.

Wang, W. S.-Y. (1976). Language change. *Annals of the New York Academy of Sciences*, *280*(1), 61–72.

Wang, W. S.-Y., Lehiste, I., Chuang, C.-K., & Darnovsky, N. (1976). Perception of vowel duration. *The Journal of the Acoustical Society of America*, *60*(S1), S92–S92.

Ward, P., & Sturt, P. (2007). Linguistic focus and memory: An eye movement study. *Memory & Cognition*, *35*(1), 73–86.

Wasz-Hockert, O. (1968). The infant cry: A spectrographic and auditory analysis. *Clinics in Development Medicine*, *29*.

Wen, Z., Mota, M. B., & McNeill, A. (2015). *Working memory in second language acquisition and processing* (Multilingual Matters, Vol. 87). Bristol Buffalo Toronto.

Werner, E., Dawson, G., Osterling, J., & Dinno, N. (2000). Brief report: Recognition of autism spectrum disorder before one year of age: A retrospective study based on home videotapes. *Journal of Autism and Developmental Disorders*, *30*(2), 157–162.

Wichmann, A., House, J., & Rietveld, T. (1997). Peak displacement and topic structure. In *Intonation: Theory, models and applications*.

Wickens, T. D. (2002). *Elementary signal detection theory*. Oxford University Press, USA.

Wildgruber, D., Ackermann, H., Kreifelts, B., & Ethofer, T. (2006). Cerebral processing of linguistic and emotional prosody: fMRI studies. *Progress in Brain Research*, *156*, 249–268.

Wildgruber, D., Hertrich, I., Riecker, A., Erb, M., Anders, S., Grodd, W., & Ackermann, H. (2004). Distinct frontal regions subserve evaluation of linguistic and emotional aspects of speech intonation. *Cerebral Cortex*, *14*(12), 1384–1389.

Willems, N. (1982). *English intonation from a Dutch point of view: An experimental-phonetic investigation of English intonation produced by Dutch native speakers* (Vol. 1). Walter de Gruyter.

Williams, J. N. (2013). Working memory and SLA. In *The Routledge handbook of second language acquisition* (Routledge, pp. 445–459).

Winston, M. E., Chaffin, R., & Herrmann, D. (1987). A taxonomy of part-whole relations. *Cognitive Science*, *11*(4), 417–444.

Wong, P. C. (2002). Hemispheric specialization of linguistic pitch patterns. *Brain Research Bulletin*, *59*(2), 83–95.

Wong, P. C., Parsons, L. M., Martinez, M., & Diehl, R. L. (2004). The role of the insular cortex in pitch pattern perception: the effect of linguistic contexts. *Journal of Neuroscience*, *24*(41), 9153–9160.

Wunderlich, D. (1988). Der Ton macht die Melodie–Zur Phonologie der Intonation des Deutschen. *Intonationsforschungen. Tübingen: Niemeyer*, 1–40.

Xu, Y., Gandour, J. T., & Francis, A. L. (2006). Effects of language experience and stimulus complexity on the categorical perception of pitch direction. *The Journal of the Acoustical Society of America*, *120*(2), 1063–1074.

Xu, Y., Gandour, J., Talavage, T., Wong, D., Dzemidzic, M., Tong, Y., … Lowe, M. (2006). Activation of the left planum temporale in pitch processing is shaped by language experience. *Human Brain Mapping*, *27*(2), 173–183.

Zárate-Sández, G. (2015). *Perception and production of intonation among English-Spanish bilingual speakers at different proficiency levels*. Georgetown University.

Zatorre, R. J., Belin, P., & Penhune, V. B. (2002). Structure and function of auditory cortex: music and speech. *Trends in Cognitive Sciences*, *6*(1), 37–46.

Ziatas, K., Durkin, K., & Pratt, C. (2003). Differences in assertive speech acts produced by children with autism, Asperger syndrome, specific language impairment, and normal development. *Development and Psychopathology*, *15*(1), 73–94.

Appendix

Screening Questionnaire

(Administered in Italian, see English translation below)

Informazioni personali

1. Email controllata di frequente (verrai contattato/a attraverso questo indirizzo es. mariox.rossi@mit.it)
2. Sesso:
3. Fascia di età: (Contrassegna solo un ovale)
    - Tra i 18 e i 35 anni
    - Tra i 36 e 40 anni
    - Tra i 41 e 50 anni
    - Tra i 51 e 60 anni
    - Più di 61 anni
4. Ti sono stati diagnosticati problemi di udito? SI NO (Contrassegna solo un ovale)
5. Sei destrimano? SI NO (Contrassegna solo un ovale)

Lingua Nativa

6. Indica la tua lingua nativa (lingua a cui sei stato esposto dalla nascita e che hai acquisito nella prima infanzia - es. Italiano)
7. Lingua nativa della madre: *
8. Lingua nativa del padre: *
9. Lingua correntemente parlata con i famigliari del nucleo ristretto (padre, madre, fratelli e sorelle)
10. Durante l'infanzia hai acquisito più di una lingua nativa? SI NO (Contrassegna solo un ovale)
11. Comunichi con i membri del tuo nucleo famigliare ristretto (madre, padre, fratelli e sorelle) attraverso più di una lingua nativa? Se sì indica quali, altrimenti scrivi "no".

Dialetti o Lingue minoritarie riconosciute in Italia

12. Indica il NOME del dialetto (o lingua minoritaria riconosciuta) che conosci meglio (es. Trentino - Romano - Salentino - Siciliano - Lingua Sarda)
13. Indica in quale AREA italiana si parla il dialetto (o lingua minoritaria riconosciuta) che hai indicato alla domanda precedente. (es. "Trentino - Trentino" / "Romano - Roma" / "Salentino - Salento" /"Lingua Sarda -
Sardegna" (indicane solo 1))
14. La tua competenza linguistica del dialetto (o lingua minoritaria riconosciuta) che hai indicato è: *
Contrassegna solo un ovale.
    - NULLA (il dialetto, o lingua minoritaria riconosciuta, equivale per te a una lingua straniera che non conosci)
    - PASSIVA E LIMITATA (comprendi vagamente il senso generale di una conversazione, non conosci la maggior parte delle parole che vengono utilizzate; non lo usi spontaneamente quasi mai;)
    - PASSIVA E AMPIA (sei in grado di comprendere il dialetto, o lingua minoritaria riconosciuta, di una specifica area, ma non lo hai mai utilizzato attivamente per comunicare, sei stato esposto durante l'infanzia a stimoli e conversazioni in dialetto o lingua minoritaria riconosciuta, lo hai sentito parlare da nonni e/o altri famigliari, ma non lo hai mai utilizzato come lingua principale con i tuoi genitori e membri del tuo nucleo stretto famigliare - genitori, fratelli e sorelle;)
    - ATTIVA E LIMITATA(sei in grado di comprendere senza problemi il dialetto, o lingua minoritaria riconosciuta, di un'area specifica, sei in grado di produrre tu stesso/a attivamente piccoli enunciati semplici; la tua competenza attiva riguardo alla grammatica e vocabolario è basica ma solida;

solitamente ti viene spontaneo utilizzare il dialetto o la lingua minoritaria riconosciuta in contesti in cui ti senti rilassato, con persone che sai lo utilizzano allo stesso tuo modo;)

- ATTIVA E AMPIA (sei in grado di comprendere e produrre enunciati lunghi e complessi; la tua competenza linguistica si avvicina, o è pari, a quella nella lingua ufficiale; utilizzi il dialetto o la lingua minoritaria riconosciuta con i famigliari più stretti, con gli amici, e rappresenta per te la lingua della tua comunità.)

15. Nella tua città/area di provenienza in Italia, il personale responsabile di esercizi commerciali e servizi che si relazionano con il pubblico (agli uffici postali, nelle banche, dentro i negozi) utilizza generalmente il dialetto o la lingua minoritaria riconosciuta quando si relaziona con i clienti? SI NO (Contrassegna solo un ovale)
16. In Italia, ti rivolgi tu stesso spontaneamente al personale presente in esercizi commerciali e servizi ai cittadini (negozi, uffici, banche, scuole, luoghi di lavoro) nel dialetto, o lingua minoritaria riconosciuta, che conosci meglio? SI NO (Contrassegna solo un ovale.)

17. Il tuo nucleo famigliare ristretto (madre, padre, fratelli e sorelle) utilizza il dialetto o lingua minoritaria riconosciuta, che hai precedentemente indicato, nella vita quotidiana? SI NO
Contrassegna solo un ovale.

18. In Italia, nel tuo cerchio stretto di amicizie (persone che conosci da tanto tempo con cui dialoghi in contesto informale) come viene utilizzato il dialetto o lingua minoritaria riconosciuta che hai precedentemente indicato? (Contrassegna solo un ovale.)

- Non viene utilizzato/a, la conversazione avviene normalmente solo in lingua ufficiale
- In modo MARGINALE rispetto alla lingua ufficiale (il dialetto o lingua minoritaria riconosciuta è utilizzato solo in precisi e ristretti contesti per comunicare con più enfasi uno stato d'animo, per comunicare sarcasmo e ironia in contesti scherzosi, per esprimere "saggezza popolare" e tradizioni del luogo come proverbi e aneddoti conosciuti dalla maggiorparte delle persone dell'area)
- In modo PARALLELO rispetto alla lingua ufficiale (dialetto o lingua minoritaria riconosciuta e lingua ufficiale vengono entrambi utilizzati, l'uso è alternato, le conversazioni possono iniziare in lingua ufficiale e proseguire in dialetto o lingua minoritaria riconosciuta senza che i parlanti percepiscano variazioni nette)
- In modo PREFERENZIALE rispetto alla lingua ufficiale (la maggioranza delle persone preferisce comunicare in dialetto o lingua minoritaria riconosciuta rispetto alla lingua ufficiale, la maggior parte delle conversazioni avviene in dialetto o lingua minoritaria riconosciuta)

Lingue Straniere
19. Lingua INGLESE: quale livello di competenza hai negli ambiti: Ascolto, Lettura, Interazione, Produzione Orale, Scritto.

- Non ho alcuna competenza A1 A2 B1 B2 C1 C2
- Ascolto
- Lettura
- Interazione
- Produzione orale
- Scritto

20. Lingua FRANCESE: quale livello di competenza hai negli ambiti: Ascolto, Lettura, Interazione, Produzione Orale, Scritto. *

- Non ho alcuna competenza A1 A2 B1 B2 C1 C2
- Ascolto

- Lettura
- Interazione
- Produzione orale
- Scritto

21. Lingua TEDESCA: quale livello di competenza hai negli ambiti: Ascolto, Lettura, Interazione, Produzione Orale, Scritto. *
    - Non ho alcuna competenza A1 A2 B1 B2 C1 C2
    - Ascolto
    - Lettura
    - Interazione
    - Produzione orale
    - Scritto

22. Lingua SPAGNOLA: quale livello di competenza hai negli ambiti: Ascolto, Lettura, Interazione, Produzione Orale, Scritto. *
    - Non ho alcuna competenza A1 A2 B1 B2 C1 C2
    - Ascolto
    - Lettura
    - Interazione
    - Produzione orale
    - Scritto

23. Se hai competenze in una lingua non indicata qui, per favore indicala e indica il livello di competenza in ogni ambito come nelle griglie precedenti.

Competenze musicali

24. Suoni correntemente e agevolmente uno strumento musicale? SI NO (Contrassegna solo un ovale.)

25. Escludendo la scuola dell'obbligo, hai mai seguito un corso con istruttori specializzati per imparare a suonare uno strumento musicale? SI NO (Contrassegna solo un ovale)

26. Se Sì, per quanti anni hai seguito un corso con istruttori specializzati?

27. Segui, o hai seguito in passato, corsi di canto con un istruttore specializzato? SI NO (Contrassegna solo un ovale.)

28. Se Sì, per quanti anni hai seguito, o da quanti anni segui corsi specializzati di canto? *

29. Hai un diploma riconosciuto che attesti le tue competenze musicali?
30. Osservando l'immagine in Figura 1, se riconosci il brano, scrivi il titolo nel box risposte. Altrimenti scrivi "no" *
Figura 1

31. Riconoscere il brano in Figura 1 per te è stato: * Contrassegna solo un ovale.
    - Estremamente Semplice
    - Abbastanza Semplice
    - Leggermente Difficoltoso
    - Estremamente Difficoltoso
    - Non ho riconosciuto il brano

Lingua Tedesca
32. Sei in possesso (o sarai in possesso entro dicembre 2016) di un attestato rilasciato da un ente riconosciuto di livello B2 o superiore (C1 o C2)?

33. Segui attualmente un corso volto al miglioramento delle tue capacità linguistiche (o al lavoro,o all'università, o nel tempo libero?) SI NO (Contrassegna solo un ovale)

Lingua Tedesca - Studio (a Scuola, nel Tempo Libero, all'Università)
34. Per quanti anni hai studiato tedesco in Italia all'interno del sistema scolastico italiano (scuole elementari, medie, e superiori)?

- Non ho mai studiato tedesco a scuola
- 3 anni alle scuole medie
- 3 anni alle scuole medie e 3 anni alle superiori
- 3 anni alle scuole medie e 5 anni alle superiori
- 3 anni alle superiori
- 5 anni alle superiori
- Altro:

35. All'interno del tuo percorso scolastico (scuole elementari, medie e superiori) ritieni che il numero di ore di lezione con un insegnante di madrelingua tedesca sia stato: *

- Non ho mai studiato tedesco a scuola
- Non ho mai fatto lezione con un insegnante di madrelingua tedesca a scuola, ho avuto
- solo insegnanti di madrelingua italiana.
- Inferiore al numero di ore con un insegnante di tedesco di madrelingua italiana
- Pari al numero di ore con un insegnante di tedesco di madrelingua italiana
- Superiore al numero di ore con un insegnante di tedesco di madrelingua italiana
- Ho avuto solo insegnanti di tedesco di madrelingua tedesca a scuola

36. All'interno del tuo percorso scolastico (scuole elementari, medie, superiori) erano previste ore di lezione e pratica unicamente dedicate ad attività volte a migliorare la conversazione e l'uso della lingua parlata?

- Non ho mai studiato tedesco a scuola
- No - ho esercitato solo la lingua scritta
- Si - SOLO con insegnanti di madrelingua ITALIANA
- Si - SOLO con insegnanti di madrelingua TEDESCA
- Sì - SIA con insegnanti di madrelingua ITALIANA SIA con insegnanti di madrelingua TEDESCA

37. Corsi di lingua nel tempo libero.

- Per quanto tempo hai frequentato un corso di lingua tedesca nel tempo libero?
- Non ho mai frequentato un corso di lingua nel tempo libero
- Da 0 a 6 mesi
- Da 7 mesi a 11 mesi
- Un anno
- Da 13 mesi a 3 anni
- Da 4 anni a 7 anni
- Più di 8 anni

38. Nei corsi di lingua che hai frequentato nel tempo libero ritieni che il numero di ore di lezione con un insegnante di madrelingua tedesca sia stato:

- Non ho mai frequentato corsi di lingua tedesca nel tempo libero
- Non ho avuto insegnanti di madrelingua tedesca nei miei corsi di lingua
- Inferiore al numero di ore con un insegnante NON di madrelingua tedesca
- Pari al numero di ore con un insegnante NON di madrelingua tedesca
- Superiore al numero di ore con un insegnante NON di madrelingua tedesca
- Ho frequentato solo corsi con un insegnante di madrelingua tedesca

39. All'interno del corso di lingua che frequenti/ hai frequentato nel tempo libero sono/erano previste ore di lezione e pratica unicamente dedicate ad attività volte a migliorare la conversazione e l'uso della lingua parlata?
    - Non ho mai frequenato corsi di lingua tedesca nel tempo libero
    - No - Mi sono esercitato solo nella lingua scritta
    - Si - SOLO con insegnanti NON di madrelingua tedesca
    - Si - SOLO con insegnanti di madrelingua TEDESCA
    - Sì - SIA con insegnanti di madrelingua TEDESCA SIA con insegnanti nativi di un'ALTRA LINGUA

40. Università:
    - Non frequento/ Non ho frequentato l'università
    - Frequento o ho frequentato un corso di laurea la cui lingua ufficiale NON è il tedesco (es. lingua ufficiale del corso: italiano o inglese)
    - Frequento o ho frequentato un corso di laurea in un'area di lingua tedesca e il tedesco è lingua ufficiale del corso (Germania, Austria, Svizzera, Alto-Adige, Belgio, Liechtenstein, Lussemburgo)

41. Frequenti o hai frequentato all'università un corso di laurea in cui lo studio della lingua tedesca è centrale e prioritario rispetto ad altri campi di studio?
    - Non frequento/Non ho frequentato l'università
    - No
    - Sì - Corsi di laurea in lingue con tedesco come prima o seconda lingua
    - Si - Corsi di laurea per interpreti e traduttori con tedesco come prima o seconda lingua
    - Si - Corsi di laurea svolti con tedesco come lingua ufficiale di tutto il corso
    - Altro:

42. Per quanti anni hai studiato tedesco (o in tedesco) all'università?
    - Non ho mai studiato tedesco all'università
    - 1
    - 2
    - 3
    - 4
    - 5
    - Altro:

43. All'interno del tuo percorso universitario ritieni che il numero di ore di lezione con un insegnante di madrelingua tedesca sia stato:
    - Non frequento/non ho frequentato l'università
    - Non ho studiato tedesco all'università
    - Non ho frequentato corsi di tedesco con insegnanti di madrelingua tedesca, ho avuto
    - insegnanti solo di madrelingua italiana o nativi di un'altra lingua
    - Inferiore al numero di ore di lezione con un insegnante di tedesco di madrelingua italiana
    - Pari al numero di ore di lezione con un insegnante di tedesco di madrelingua italiana
    - Superiore al numero di ore con un insegnante di tedesco di madrelingua italiana

- I miei insegnanti di tedesco sono/sono stati solo di madrelingua tedesca
- Il mio intero corso di laurea (non solamente corso di lingua) si svolge/si è svolto in tedesco

44. All'interno del tuo percorso universitario sono/erano previste ore di lezione e pratica unicamente dedicate ad attività volte a migliorare la conversazione e l'uso della lingua parlata?
- Non frequento/non ho frequentato l'università
- Non ho studiato tedesco all'università
- Non sono/erano previste attività dedicate al miglioramento della lingua parlata ma l'intero
- corso di laurea si svolge/si è svolto in tedesco
- No - ho esercitato solo la lingua scritta
- Si - SOLO con insegnanti NON di madrelingua tedesca
- Si - SOLO con insegnanti di madrelingua tedesca
- Sì - SIA con insegnanti di madrelingua TEDESCA SIA con insegnanti nativi di un'ALTRA LINGUA

<u>Lingua Tedesca - Permanenza in un' Area di Lingua Tedesca</u>

45. La prima volta in cui ti sei trasferito per un tempo consecutivo di almeno di un mese in un'area di lingua tedesca (Germania, Austria, Svizzera, Alto-Adige, Belgio, Liechtenstein, Lussemburgo) quanti anni avevi?

46. Nell'area di lingua tedesca (Germania, Austria, Svizzera, Alto-Adige, Belgio, Liechtenstein, Lussemburgo) in cui attualmente vivi, sono presenti altri membri del tuo nucleo famigliare allargato d'origine (genitori, fratelli o sorelle, zii, cugini, nonni)? Sì
- No
- Al momento della compilazione del questionario, non sto vivendo in un'area di lingua tedesca (Germania, Austria, Svizzera, Alto-Adige, Belgio, Liechtenstein o Lussemburgo)

47. Nell'area di lingua tedesca (Germania, Austria, Svizzera, Alto-Adige, Belgio, Liechtenstein, Lussemburgo) in cui attualmente vivi, quale lingua o dialetto utilizzi maggiormente per conversare con gli amici nei momenti di svago?
- Inglese
- Tedesco
- Italiano
- Al momento della compilazione del questionario. non sto vivendo in un'area di lingua tedesca (Germania, Austria, Svizzera, Alto-Adige, Belgio, Liechtenstein, o Lussemburgo)
- Altro:

48. Per quanto tempo hai vissuto continuativamente in un paese di lingua tedesca (Germania, Austria, Svizzera, Alto-Adige, Belgio, Liechtenstein, Lussemburgo)?
- Non ho mai vissuto in un paese di lingua tedesca
- Da 0 a 2 mesi
- Da 3 a 6 mesi
- Da 7 a 11 mesi
- Un anno
- Da 13 mesi a 3 anni
- Da 4 a 7 anni
- Da 8 a 15 anni
- Più di 16 anni
- Altro:

49. Da quanto tempo stai vivendo continuativamente ora (al momento della compilazione del questionario) in un paese di lingua tedesca (Germania, Austria, Svizzera, Alto-Adige, Belgio, Liechtenstein, Lussemburgo)?

- Non sono al momento in un paese di lingua tedesca
- Da 0 a 2 mesi
- Da 3 a 6 mesi
- Da 7 a 11 mesi
- Un anno
- Da 13 mesi a 3 anni
- Da 4 a 7 anni
- Da 8 a 15 anni
- Più di 16 anni

50. Ti trovi al momento in un'area di lingua tedesca (Germania, Austria, Svizzera, Alto-Adige, Belgio, Liechtenstein, Lussemburgo) per:

- Formazione Scuola - Lavoro (Ausbildung)
- Studio
- Dottorato
- Lavoro
- Svago
- Non mi trovo al momento in un'area di lingua tedesca

51. Se sei un lavoratore in un'area di lingua tedesca (Germania, Austria, Svizzera, Alto-Adige, Belgio, Liechtenstein, Lussemburgo), per svolgere la tua attività lavorativa devi NECESSARIAMENTE interagire con i colleghi in tedesco?

- SI
- NO
- Non sono un lavoratore in un'area di lingua tedesca

52. Quando NON sei in un'area di lingua tedesca (Germania, Austria, Svizzera, Alto-Adige, Belgio, Liechtenstein, Lussemburgo) interagisci comunque in forma orale con parlanti nativi del tedesco? *

- Si
- No
- Altro:

Lingua Tedesca - Materiale audiovisivo

53. Indica quante ORE guardi mediamente in UNA SETTIMANA materiale audiovisivo IN TEDESCO. Indica le ore in una settimana in cifre.

54. Di solito quando fruisci di contenuti e materiali audiovisivi IN TEDESCO:

- Non utilizzi alcun tipo di sottotitolo
- Utilizzi sottotitoli in tedesco
- Utilizzi sottotitoli in italiano
- Non fruisco mai di contenuti e materiali audiovisivi IN TEDESCO

55. Quando fruisci di contenuti e materiali audiovisivi:

- Non fruisco mai di contenuti e materiali audiovisivi
- Comprendi un piccola parte dei contenuti, spesso perdi il filo conduttore del discorso, lo trovi stancante e non ti diverte. Non la consideri un'attività ricreativa.

- Comprendi generalmente il senso del discorso ma è un'attività che puoi fare per periodi limitati, ti diverte ma alla fine ti devi riposare un po'. E' un attività leggermente ricreativa.
- Comprendi quasi tutto e puoi farlo per periodi lunghi. Ogni tanto hai qualche difficoltà a capire contenuti particolari come scherzi, battute, parole che non usi di frequente. E un'attività tutto sommato ricreativa.
- Comprendi tutti gli scambi linguistici, anche le battute umoristiche. Puoi farlo per tanto tempo. E' un'attività ricreativa. Non noti la diversa provenienza di tutti gli interlocutori solo dal modo in cui parlano.
- Comprendi tutti i contenuti e sei in grado di identificare vari accenti e la provenienza dei personaggi che interagiscono nel programma. E' un'attività ricreativa.

Lingua Tedesca - Programmi radiofonici

56. Indica quante ORE ascolti mediamente in UNA SETTIMANA programmi radiofonici attraverso qualunque supporto audio o piattaforma, IN TEDESCO. *
Indica le ore in una settimana in cifre.

57. Riconoscere e comprendere IL PARLATO di programmi radiofonici IN TEDESCO per te: *

- Contrassegna solo un ovale.
- Estremamente semplice
- Abbastanza semplice
- Leggermente difficoltoso
- Estremamente difficoltoso
- Non riesco a comprendere il parlato di programmi radiofonici in tedesco

Screening Questionnaire (English translation)

Personal Information

1)Email you frequently control (you will be contacted through this email address)
2)Sex
3)To what age group do you belong?

- -between 18 and 35
- 36 and 40
- 41 and 50
- 51 and 60
- More than 61

4) Have you been diagnosed with hearing problems? Yes no
5) Are you Right-handed? Yes No

Native Language

6) Please state your native language (first language you acquired in infancy) ___________
7) Native Language of your mother (e.g. Italian) ________________________
8) Native Language of your father (e.g. French) ____________________________
9) Which language do you currently speak with the members of your restricted nucleus of family (father, mother, siblings) ________________
10) During infancy did you acquire more than one native language? Yes No.
11) Do you communicate with the members of your restricted nucleus of family (mother, father, siblings) through more than one native language? Yes No. If you communicate through more than one language please state which ones

Use of Dialects or of minority Languages

12) Which dialect do you know better? (state the name or the area in which is spoken)
13) Please state in which AREA of Italy you speak the dialect you have indicated in the previous question

14) The linguistic competence you have in the dialect you know better is:

- NONE
- PASSIVE AND LIMITED (you understand roughly the general meaning of a conversation, you don't know the majority of the words, you almost never use it spontaneously)
- PASSIVE AND BROAD (you are able to understand a dialect of a specific area, but you never actively used it to communicate, you have been exposed to it in infancy by family members as grandparents or other relatives, but you never used in your restricted family nucleus – parents, siblings)
- ACTIVE AND LIMITED (you have no problems understanding the dialect of a specific area, you yourself are able to actively produce small sentences; your active knowledge about grammar and vocabulary is basic but solid; you usually use the dialect in informal context with people you know use it like you do)
- ACTIVE AND BROAD (you fully understand and produce complex sentences in dialect; your linguistic competence is close to, or equal to the official language; you use language with your closest family members, friends, and is for you the language of your community)

15) In the area city where you come from in Italy, the people working in commercial areas and services for the public (at the post office, at the bank, in shops) generally use dialect when talking to the clients? Yes No

16) In Italy, do you spontaneously use dialect to talk to the people working in commercial area and services for the citizens (shops, offices, banks, schools, job's places)? Yes No

17) Does your restricted family nucleus (mother father siblings) use dialect (the one you stated above) in everyday life? Yes No

18) In Italy, in your restricted circle of friends (people you know from a very long time with whom you interact in an informal context) how is the dialect used?

- In a PREFERENTIAL way compared to the official language (the majority of the people prefer dialect compared to the official language, the majority of the conversations happen in in dialect)

- In a PARALLEL way compare to the official language (dialect and official language are both used, the use is alternated, conversations can start in the official language and end in dialect and the speakers don't perceive any sharp variation)

- in a MARGINAL way compared to the official language (the dialect is used just in specific and restricted context to communicate with more emphasis an emotional state, to communicate irony and sarcasm in goliardic contexts, to express "traditional popular knowledge" like way to say and anecdotes that the majority of the people of the area share)

- It's not used, conversation happens only in the official language

Foreign Languages

19) English: please state your CEFR level in each area:

	No competence	A1	A2	B1	B2	C1	C2
Listening	○	○	○	○	○	○	○
Reading	○	○	○	○	○	○	○
Interaction	◉	○	○	○	○	○	○
Oral Production	○	○	○	○	○	○	○
Written Production	○	○	○	○	○	○	○

20) French
21) German
22) Spanish
23) If you have any linguistic competence in a language not listed above, please state it here.

Musical Competence

24) Do you currently and fluently play a musical instrument? Yes No

25) Excluding the mandatory school (#elementary, middle, part of high school) have you ever taken classes with a qualified instructor to learn to play a musical instrument?

26) If Yes, for how many years did you follow classes with qualified instructors?

27) Are you taking, or have you taken in the past, singing classes with a qualified instructor? Yes No

28) If Yes, how many years have you followed, or since how many years do you follow singing classes with a qualified instructor?

29) Do you own a recognized musical diploma, of which level?

Figure 1

30) Looking at the picture in Figure 1, if you recognize the musical piece from which it comes from, write the name in the answer box. Otherwise write "no".__________________________

31) Recognizing the musical piece in figure 1 for you it has been:

- extremely easy
- fairly easy
- slightly difficult
- very difficult
- I didn't recognize the musical piece

German Language (GL)

32) Do you hold (or will you be holding by December 2016) a certificate from a recognized institute of B2 Level or above (C1 or C2)?

33) GL: Are you currently following a class with the goal to improve your linguistic skills (at work, or at university, o in your free time)? Yes No

34) GL: How many years did you study German in Italy within the Italian scholastic system (elementary, middle, high schools)? If you studied German at school but you don't find your condition please state it in "other".

- 3 years in middle school
- 3 y in middle school and 3 y in high school
- 3 y in middle school and 5 y in high school
- 3 y in high school
- 5 y in high school
- I have never studied German at school
- Other ____________________________

35) GL: Within your scholastic cycle (elementary, middle, high school) do you think that the hours of class with native speaker of German as teacher has been:

- I have never studied German in school
- I have never had class with German native speaker teacher, I just had teachers native speakers of Italian
- less than the hours with an Italian native speaker as teacher

- equal to the hours with an Italian native speaker as teacher
- More than the hours with an It native speaker as teacher
- I just had native speaker of German as teachers

36) Within your school years (elementary, middle, high school) did you attend classes specifically devoted to activities to improve conversation and spoken language use?
- I have never studied German in school
- No – I exercised just the written language
- Yes – ONLY with teacher native speakers of ITALIAN
- Yes ONLY with teacher native speakers of GERMAN
- Yes – BOTH WITH teachers native speaker of ITALIAN AND teachers native speakers of GERMAN

37) GL: University

- I have never attended/I don't attend university
- I attend or have attended a degree whose official language is NOT German (Es. Official language of the degree: Italian or English)
- I attend or have attended a degree in German Speaking Area *(#from now on GSA: refers to Germany, Austria, Switzerland, Sued Tirol, Belgium, Liechtenstein and Luxemburg)* whose official language was German

38) GL: Do you attend or have you attended at university a degree in which the study of German Language is central or has a priority compared to other subjects? Specify in "Other" your situation if not described in the options.
- I don't attend/didn't attend university
- No
- Yes – Degrees in Foreign Languages with German as First or Second Language
- Yes – Degrees in Translation and Interpreting with German as First or Second Language
- Yes- Entire degree with German as official language of the course
- Other ________________________

39) GL: How many years have you studied German (or in German) at University?
- Never
- 1
- 2
- 3
- 4
- 5
- Other __________________________________

40) GL: In the years you attended University, do you think that the hours of class with native speaker of German as teacher has been:
- I don't attend/didn't attend University
- I have never studied German at University
- I have never had class with German native speaker teacher, I just had teachers native speakers of Italian or of another language;
- less than the hours with a teacher native speaker of language other than German
- equal to the hours with a teacher native speaker of language other than German
- More than the hours with a teacher native speaker of language other than German
- I just had native speaker of German as teachers

- My entire degree has been in German (not just language classes)

41) During University did you attend classes specifically devoted to activities to improve conversation and spoken language usage?
- I don't attend/didn't attend University
- I have never studied German at University
- Spoken language classes were not provided but my whole degree has been in German
- No – I exercised just the written language
- Yes – ONLY with a teacher native speaker of a language other than German
- Yes ONLY with a teacher native speaker of a language other than German
- Yes – BOTH WITH teachers native speaker of ITALIAN AND teachers native speakers of GERMAN

42) GL : Language classes in your **spare time**. How long have you attended a German class in your free time? Never
- From 0 to 6 months
- From 7 months to 11 months
- 1 year
- from 13 months to 3 years
- from 4 years to 7 years
- more than 8 years
-

43) GL : In the Language classes you attended in your free time, do you think the hours of class with a teacher who is native speaker of German have been:
- I never attended a German class in my spare time
- I never had teachers native speakers of German in my German classes
- less than the hours with a teacher who is NOT native speaker of German
- equal to the hours with a teacher who is NOT native speaker of German
- more than the hours with a teacher who is NOT native speaker of German
- I just attended classes with teachers who were native speakers of German

44) GL : In your language classes you attended in your free time, were there activities and practices in which you could just train to improve conversation and spoken language usage?
- I never attended language classes in my free time
- No I trained just the written language
- Yes – ONLY with teacher who were NOT native language speakers of German
- Yes _ ONLY with teachers who were native language speakers of GERMAN
- Yes – BOTH WITH teacher who were L1 GERMAN AND teachers who were L1 in ANOTHER LANGUAGE

Immersion in a German Speaking Area (GSA)

45) GL: The first time you move in GSA for a consecutive time of at least 1 month you were __ years old: If you have never lived in a GSA write "no". In other cases please write your age in numbers. If the first time is now at the moment of answering, please state your actual age. _____________________-

46) GL: In the GSA you are currently living in, are there other members of your original restricted family nucleus (parents, siblings) ? Yes, No, At the moment I am not living in a GSA

47) GL: In the GSA in which you are currently living, which language or dialect do you mostly use to talk to friends when you are enjoying your time? Please state just one preference. If it doesn't appear in the choices write it in "Other".

- English
- German
- Italian
- At the moment I am not in a GSA
- Other

48) GL: How long have you lived continuatively in a GSA? If you have lived in a GSA in different occasions (once for 6 months and once for 2 years) please sum them up (2 y and 6 months – select from 13 months to 3 years)

- Never
- From 0 to 2 months
- From 3 to 6 months
- From 7 to 11 months
- A year
- From 13 months to 3 years
- From 4 y to 7 y
- More than 8 y
- Other_________________

49) How long have you been living continuatively now (at the time of answering) in a GSA? If you have lived in a GSA in different occasions please consider just the period of time at the time of answering (1 y in 2002, I came back to a GSA in august 2016 – select from 3 to 6 months)

- I am not at the moment in a GSA
- from 0 to 2 months
- from 3 to 6 months
- from 7 to 11 months
- 1 y
- from 13 months to 3 years
- from 4 y to 7 y
- more than 8 y

50) At the moment you are in a GSA for:

- School/Job Training (Ausbildung)
- Study
- Ph.D.
- Work
- Leisure
- I am not in a GSA at the moment

51) GL: If you are a worker in a GSA to fully carry out your working activity do you NECESSARILY have to interact with colleagues in German? Click No if you can carry out your job in Italian or English or any other language but German. Ausbildung is considered a working activity. Ph.D. is considered a working activity. If you are a student answer "I am not a worker in a GSA".
-Yes /No/ I am not a worker in a GSA

52) GL: When you are NOT in a GSA do you anyway interact in an oral way with German native speakers? (Yes, if for example you have German friends in Italy, yes if you never stop taking classes, yes if you always do Tandem activities. Yes if from Italy you work everyday with German colleagues. Other if you have a particular condition that you find relevant but is not in the options. Yes No Other ________________

German Language: Audio-Visual Material

53) GL: How many hours per week do you watch Audio-Visual material in German TV?

54) GL: To follow Audio-Visual material in German for you is:

- I can't follow a radio-TV program or a movie in German
- Possible only with Italian subtitles
- Possible only with Italian or German subtitles
- Possible also without subtitles

55) GL: When you follow Audio-Visual material in German you:

- I don't follow Audio-Visual material in German
- I understand a small part of the content, often you lose the conducting theme of the conversation, you find the activity tiring and you don't have fun. You don't consider it a recreational activity.
- I generally understand the global meaning of what it has been said but it is an activity you can do for delimited periods of time, you have fun by it but at the end you need some rest. It is slightly a recreational activity.
- I understand almost everything and I can do it for long periods of time. Sometimes you have some difficulties to understand jokes, or words you don't frequently use. Overall is a recreational activity.
- I understand everything, also sarcasm and jokes. You can do it for a long time. It is a recreational activity. You don't notice different accents and where the speakers come from just by hearing them talking.
- I understand all the contents and you identify different accents and where the speakers come from. It is a fun activity.

German Language: Hours of Radio

56) GL: Do you listen to radio in German?

- Never
- Rarely
- Occasionally
- Often
- Always

57) To understand the spoken language used in Radio programs for you is:

- I can't understand spoken language when listening to Radio
- Extremely difficult
- Rather difficult
- Rather easy
- Extremely easy

Sentences used in the EEG experiment in Chapter 4

1.Tilmann kauft sich eine Flasche Wein. Als erstes riecht er am Korken.
2. Die Putzfrau macht ein Bett. Als erstes schüttelt sie das Kissen.
3. Robert sieht in der Buchhandlung ein Buch. Zuerst liest er den Titel.
4. Helga besorgt sich ein Auto. Zuerst kontrolliert sie die Räder.
5. Manfred bringt eine Tasche in das Zimmer. Dann setzt er sich an den Schreibtisch.
6. Christian verletzt sich an der Hand. Deshalb verbindet er den Daumen.
7. Josef mietet sich eine Wohnung. Zuerst streicht er die Wände.
8. Klara besichtigt eine Stadt. Dabei fotografiert sie das Rathaus.
9. Maria sucht einen neuen Computer. Dabei achtet sie auf den Bildschirm.
10. Tim hängt ein Bild auf. Allerdings hasst er den Rahmen.
11. Doris geht mit Tim in einen Park. Dort fotografiert er die Bäume.
12. Hanna läuft durch einen Supermarkt. Endlich findet sie die Kasse.
13. Julia zeigt dem Polizisten ihren Reisepass. Dabei lacht er über das Passbild.
14. Fred dreht sich eine Zigarette. Allerdings vergisst er den Filter.
15. Constantin kauft seiner Frau ein Armband. Als erstes kontrolliert er die Perlen.
16. Daniela beobachtet im Hafen ein Boot. Überraschenderweise gefällt ihr das Segel.
17. Lena mag gepflegte Finger. Deswegen lackiert sie ihre Nägel.
18. Noel arbeitet in einem Hotel. Dort putzt er die Zimmer.
19. Die Lehrerin ist im Klassenzimmer. Dort steht sie vor der Tafel.
20. Petra fährt ein neues Mountainbike. Als erstes kontrolliert sie die Bremse.
21. Markus wartet vor einem Kino. Dort steht er beim Eingang.
22. Jasmin sitzt in einer Kirche. Dort bewundert sie die Kuppel.
23. Laura läuft den ganzen Sonntag durch ein Museum. Gegen Abend sucht sie den Ausgang.
24. Jens arbeitet als Schauspieler in einem Theater. Jeden Tag steht er auf der Bühne.
25. Felix reinigt ein Klavier. Als erstes putzt er die Tasten.
26. Paula verbringt viel Zeit in der Bibliothek. Deswegen kennt sie die Bücher.
27. Jacob bestellt zum Frühstück ein Ei. Dann klopft er auf die Schale.
28. Marco eröffnet ein Restaurant. Davor renoviert er die Küche.
29. Verena redet mit ihrem Opa im Wohnzimmer. Dort sitzt er im Sessel.
30. Irina hat ein schönes Gesicht. Vor allem mag sie ihre Nase.
31. Rainer zeichnet gerne Gesichter. Dabei beginnt er mit den Augen.
32. Gitta fotografiert einen Berg. Zuerst fokussiert sie den Gipfel.
33. Laura hat ein altes Telefon. Am schönsten findet sie den Hörer.
34. Daniel putzt eine Küche. Zuerst leert er den Kühlschrank.
35. Rita schreibt mit einem Bleistift. Leider zerbricht sie die Mine.
36. Siegfried kauft sich ein Glas Marmelade. Allerdings beschädigt er den Deckel.
37. Barbara repariert eine Lampe. Zunächst ersetzt sie die Birne.
38. Lukas isst eine Wassermelone. Dabei entfernt er die Kerne.
39. Karl möchte in seine Wohnung. Deshalb öffnet er die Haustür.
40. Sophia studiert die Anatomie eines menschlichen Körpers. Aufmerksam betrachtet sie die Muskeln.
41. Constantin putzt seine Brille. Zuerst poliert er die Gläser.
42. Klaus kontrolliert seinen Wagen. Dann repariert er den Motor.
43. Martina schließt das Fenster. Dabei berührt sie die Scheibe.

44. Helga betrachtet einen Baum. Dann zeichnet sie die Blätter.
45. Peter kümmert sich um einen Garten. Als letztes pflanzt er die Blumen.
46. Lisa sucht ihren Zug im Hauptbahnhof. Dabei verwechselt sie die Gleise.
47. Mario hat Schmerzen im Mund. Deswegen überprüft er die Zähne.
48. Anna renoviert ein Haus. Am Wochenende streicht sie den Keller.
49. Thomas kauft Möbel für ein Schlafzimmer. Allerdings vergisst er den Nachttisch.
50. Sebastian bezieht das Bett. Als erstes nimmt er die Decke.
51. Olga setzt sich auf einen Stuhl. Dabei beschädigt sie die Lehne.
52. Rainer sichert sein Fahrrad mit einem Schloss. Leider verliert er den Schlüssel.
53. Jana kauft eine neue Gitarre. Nach einer Woche ersetzt sie die Saiten.
54. Diana spielt mit Freunden Billard. Dabei konzentriert sie sich auf die Kugeln.
55. Die Großmutter backt einen Kuchen. Danach verteilt sie die Stücke.
56. Stephan putzt ein Badezimmer. Allerdings vergisst er den Spiegel.
57. Maria versendet einen Brief. Dazu braucht sie den Umschlag.
58. Magda holt eine Banane. Dann entfernt sie die Schale.
59. Sarah macht Ordnung in ihrem Schrank. Zuerst leert sie die Fächer.
60. Der Junge malt einen Vogel. Zuerst zeichnet er die Flügel.
61. Lola klemmt sich eine Hand in der Tür ein. Dabei bricht sie sich die Finger.
62. Silvia befindet sich in einem Gewitter. Dabei hört sie den Donner.
63. Kerstin streichelt eine Katze. Dabei berührt sie die Ohren.
64. Robert sieht einen Regenbogen. Dabei bewundert er die Farben.
65. Ein Kind ist mit seiner Familie am See. Dort beobachtet es die Fische.
66. Henning holt sich eine Tasse. Vorsichtig umfasst er den Henkel.
67. Ein Soldat versucht den Fuß des Gegners zu verletzen. Dabei trifft er die Ferse.
68. Der Feuerwehrmann parkt das Feuerwehrauto. Sehr schnell holt er die Leiter.
69. Franz repariert ein Fahrrad. Dabei wechselt er den Reifen.
70. Andreas repariert ein altes Hemd. Dabei ersetzt er die Knöpfe.
71. Ingo öffnet schnell eine Tür. Beim Festhalten zerbricht er die Klinke.
72. Heiko betrachtet einen sehr großen Vogel. Dabei bewundert er den Schnabel.
73. Robin Hood schießt einen Pfeil ab. Nach einer Weile findet er die Spitze.
74. Verena sieht einen Elefanten. Plötzlich bewegt er den Rüssel.
75. Stephan renoviert ein Badezimmer. Zunächst montiert er die Dusche.
76. Sabine findet einen alten Schuh. Dann repariert sie die Sohle.
77. Dieter greift nach einer Zahnpastatube. Beim Öffnen entfernt er die Kappe.
78. Isabel arbeitet im Aquarium mit einem Delphin. Zum Gruß bewegt er die Flosse.
79. Carla hat Angst vor der Wespe. Natürlich fürchtet sie den Stachel.
80. Kurt kauft eine Kamera. Zu Hause poliert er die Linse.
81. Peter betrachtet eine Standuhr. Mit den Augen folgt er dem Pendel.
82. Der Zahnarzt zieht einen Zahn. Dabei achtet er auf die Wurzel.
83.Brigitte sieht im Zoo ein Kamel. Von weitem erkennt sie die Höcker.
84.Der Polizist zielt mit einer Pistole. Dann drückt er den Abzug.
85.Alissa kauft ein altes Fahrrad. Danach ersetzt sie den Sattel.
86.Helga entdeckt einen Schuhkarton. Dann öffnet sie den Deckel.
87.Herbert rührt in einem Topf. Dann holt er den Deckel.

88. Lisa liebt rote Rosen. Allerdings hasst sie die Dornen.
89. Nora kuschelt zu Hause mit ihrem Hund. Dabei berührt sie die Schnauze.
90. Carolin fährt mit einem Wagen. Um 5 Uhr betätigt sie die Lichter.
91. Christina repariert eine Armbanduhr. Dabei überprüft sie die Zeiger.
92. Silvia steht vor einem Kirchturm. Plötzlich hört sie die Glocken.
93. Frederick fährt auf dem Fluss mit einem Boot. Dabei bewegt er die Ruder.
94. David schenkt Lisa eine schöne Blume. Dabei zerknickt er den Stängel.
95. Der Pilot landet ein Flugzeug. Dananch verlässt er das Cockpit.
96. Knut renoviert ein altes Dach. Dabei ersetzt er die Ziegel.
97. Georg fischt am Fluss mit einer langen Angel. Zuerst präpariert er den Haken.
98. Maria schaut von oben auf viele Häuser. Dabei sieht sie die Dächer.
99. Nina kauft sich ein neues schweres Buch. Beim Öffnen streichelt sie den Einband.
100. Rebecca kauft sich eine neue Bluse. Zu Hause kürzt sie die Ärmel.
101. Hans fährt mit einem Fahrrad. Dabei benutzt er die Klingel.
102. Jennifer renoviert in ihrer Wohnung den Boden. Dabei wechselt sie die Fliesen.
103. Edith pflanzt im Garten einen Baum. Dabei berührt sie die Rinde.
104. Daniel benutzt eine Toilette. Danach betätigt er die Spülung.
105. Thorsten geht an einem Fluss entlang. Nach einer Weile sieht er die Dämme.
106. Mandy kauft Baumwolle. Als erstes kontrolliert sie die Fasern.
107. Lea spielt eine Melodie. Allerdings vergisst sie die Noten.
108. Luis zieht die Schuhe an. Dabei bindet er die Schnüre.
109. Sabine arbeitet den ganzen Tag. Dabei zählt sie die Stunden.
110. Thomas kauft sich zwei neue Messer. Dann poliert er die Griffe.
111. Lea setzt sich in ihr neues Auto. Dann umfasst sie das Lenkrad.
112. Gisela zerreißt ein Schulheft. Danach faltet sie die Blätter.
113. Mario geht ins Klassenzimmer. Dann putzt er die Tafel.
114. Tim will im Schlafzimmer das Licht anmachen. Deswegen sucht er den Schalter.
115. Bob kauft einen neuen Schrank. Als erstes kontrolliert er die Türen.
116. Mandy sucht in der Menge nach einem bekannten Gesicht. Dann erkennt sie die Haare.
117. Christiane kauft sich einen Trekkingrucksack. Dabei kontrolliert sie die Träger.
118. Ein Polizist wäscht eine Uniform. Danach bügelt er die Hose.
119. Marco kauft einen gebrauchten Laptop. Allerdings fehlt das Netzteil.
120. Julia füttert eine Schildkröte. Danach streichelt sie den Panzer.
121. Jens hilft den Schauspielern auf der Bühne. Dabei öffnet er den Vorhang.
122. Der Schriftsteller schreibt ein Buch. Am Ende organisiert er den Anhang.
123. Lisa spielt auf dem Spielplatz. Dabei benutzt sie die Rutsche.
124. Julian macht Ordnung auf der Festplatte eines Laptops. Dabei löscht er die Ordner.
125. Franz überprüft eine Internetseite. Dabei probiert er die Web-Links.
126. Antonio reist durch Europa. Dabei schreibt er über die Länder.
127. Anna verschließt eine Tür. Dabei zerbricht sie den Riegel.)
128. Giovanni macht einen Kaffee. Als erstes nimmt er den Filter.
129. Daniel montiert ein Spülbecken. Unten präpariert er die Leitung)
130. Stephan rennt durch ein Kornfeld. Dabei berührt er die Ähren.
131. Hannah wirft den Abfall in die Tonne. Dabei zerreißt sie den Müllsack.

132. Verena zündet eine Kerze an. Dabei sieht sie die Flamme.
133. Milena macht einen frischen Orangensaft. Dabei sammelt sie das Fruchtfleisch.
134. Ein Torero kämpft in der Arena gegen einen Stier. Dabei fixiert er die Hörner.
135. Lucas kauft in Deutschland ein elektronisches Gerät. In Italien wechselt er den Stecker.
136. Peter schaut eine Baseballausrüstung an. Dabei probiert er den Schläger.
137. Anna schminkt ihr Gesicht. Dabei bemalt sie die Lippen.
138. Paul arbeitet auf dem Fußballplatz. Deswegen achtet er auf die Tore.
139. Der Kinderarzt bittet Jana den Mund aufzumachen. Dann zeigt sie die Zunge.
140. Valentina zeichnet eine Blume. Danach bemalt sie die Blüte.
141. Tim arbeitet bei einer Bank. Dort sitzt er hinter dem Schalter.
142. Milena macht sich ein Rührei. Dabei verquirlt sie das Dotter.
143. Louise macht sich einen Cappuccino. Zuerst röstet sie die Bohnen.
144. Albert repariert ein Telefon. Dabei untersucht er das Kabel.