Designing new experiences of music making

A thesis submitted to the University of Trento for the degree of Doctor of Philosophy in the Department of Information Engineering and Computer Science, International Doctoral School in Information and Communication Technologies

FABIO MORREALE

26TH PH.D. CYCLE

ADVISOR
Prof. Antonella De Angeli, University of Trento (Italy)

THESIS EXAMINERS
Prof. Roberto Bresin, KTH Stockholm (Sweden)
Prof. Ernest Edmonds, University of Technology, Sydney (Australia)
Prof. Fabio Pittarello, University Ca' Foscari of Venice (Italy)

JANUARY 27TH, 2015

TABLE OF CONTENTS

ABSTRACT

Music making is among the activities that best fulfil a person's full potential, but it is also one of the most complex and exclusive: successful music making requires study and dedication, combined with a natural aptitude that only gifted individuals possess. This thesis proposes new design solutions to reproduce the human ability to make music. It offers insights to provide the general public with novel experiences of music making by exploring a different interactive metaphor. Emotions are proposed as a mediator of musical meanings: an algorithmic composer is developed to generate new music, and the player can interact with the composition, controlling the desired levels of the composition's emotional character. The adequacy of this metaphor is tested with the case study of The Music Room, an interactive installation that allows visitors to influence the emotional aspect of an original classical style musical composition by means of body movements. This thesis addresses research questions and performs exploratory studies that are grounded in and contribute to different fields of research, including musical interface design, algorithmic composition, and psychology of music. The thesis presents MINUET, a conceptual framework for the design of musical interfaces, and the Music Room, an example of interactive installation based on the emotional metaphor. The Music Room was the result of a two-year iteration of design and evaluation cycles that informed an operational definition of the concept of engagement with interactive art. New methods for evaluating visitors' experience based on the integration of evidences from different user-research techniques are also presented. As regards the field of algorithmic composition, the thesis presents Robin, a rule-based algorithmic affective composer, and a study to test its validity in communicating different emotions in listeners (N=33). Valence (positive vs. negative) and arousal (high vs. low) were manipulated in a 2*2 within-subjects design. Results showed that Robin correctly communicated valence and arousal in converging conditions (high valence, high arousal and low valence, low arousal). However, in cases of diverging conditions (high valence, low arousal and low valence, high arousal), valence received neutral values. As regards the psychology of music, this thesis contributes new evidence to the on-going debate about the innate or learned nature of musical competence, defined as the ability to recognise emotion in music. Results of an experimental study framed within Russell's two-dimensional theory of emotion suggest that musical competence is not affected by training when listeners are required to evaluate arousal (dictated by variations of tempo). The evaluation of valence (dictated by the combination of tempo and mode), however, was found to be more complicated, highlighting a difference in the evaluation of musical excerpts when tempo and mode conveyed diverging emotional information. In this debate, Robin is proposed as a suitable tool for future experimental research as it allows the manipulation of individual musical factors.

STATEMENT OF CONTRIBUTION

This disclaimer is to state that the research reported in this thesis is primarily the work of the author and was undertaken as part of his doctoral research. In all the referenced paper the student is the leading author.

The work reported in Chapters 3,4, 5 and 6 has been published as follows. The content of these papers has been re-interpreted and rewritten in the thesis.

The study reported in Chapter 3 was published in parts as

• Morreale, F., Masu, R., De Angeli, A., & Fava, P. (2013). The Effect of expertise in evaluating emotions in music. In *Proceedings of the 3rd International Conference on Music & Emotion (ICME3), Jyväskylä, Finland, 11th-15th June 2013. Geoff Luck & Olivier Brabant (Eds.). ISBN 978-951-39-5250-1*. University of Jyväskylä, Department of Music.

• Morreale, F., Masu, R., & De Angeli, A. (2013). Robin: an algorithmic composer for interactive scenarios. *Proceedings of SMC*, *2013*, 10th.

The study reported in Chapter 4 was published as

• Morreale, F., De Angeli, A., Masu, R., Rota, P., & Conci, N. (2014). Collaborative creativity: The Music Room. *Personal and Ubiquitous Computing*, *18*(5), 1187-1199.

• Morreale, F., Masu, R., De Angeli, A., & Rota, P. (2013, April). The music room. In *CHI'13 Extended Abstracts on Human Factors in Computing Systems* (pp. 3099-3102). ACM.

The study reported in Chapter 5 was published in parts as Morreale, F., De Angeli, A., & O'Modhrain, S. Observations on visitors' behaviour in The Music Room. *Proc. of Practice-Based Research Workshop at NIME '14.*

The study reported in Chapter 6 was published in parts as Morreale, F., De Angeli, A., & O'Modhrain, S. (2014). Musical Interface Design: An Experience-oriented Framework. In *Proc. of NIME* (Vol. 14).

ACKNOWLEDGMENTS

I would like to thank my advisor Prof. Antonella De Angeli for her unceasing support and guidance. She has been a huge mentor for me. Thanks for encouraging my research and for allowing me to grow as a research scientist.

I am grateful to Prof. Sile O'Modhrain, for having me as a visiting researcher at the University of Michigan.

I also thank Prof. Roberto Bresin, Prof. Ernest Edmonds, and Prof. Fabio Pittarello for examining this thesis and for providing crucial feedbacks and comments.

I want to express my gratitude to all my colleagues of the interAction group at the University of Trento for their extensive support throughout my Ph.D. years. Raul Masu deserves a special mention for his generous and crucial assistance. Thanks to all my coauthors Aliaksei Miniukovich, Paolo Rota, Nicola Conci, Patrizio Fava and Maria Teresa Chietera. Special thanks go to the undergraduate students I co-supervised for their remarkable commitment, and to Costanza Vettori who helped me editing the papers. And thanks to Ivan Favalezza for his help with the layout of this thesis. Heartfelt thanks to Chiara, Boghi and my lifelong friends.

Last but not least, I am profoundly grateful to my family for the unconditional love and support.

“In the best of all possible worlds,
art would be unnecessary.
Its offer of restorative, placative therapy
would go begging a patient.
The professional specialization involved
in its making would be presumption...
The audience would be
the artist and their life would be art.”
*[**Gould, Glenn**]*

Gloud, G. (1966). The Prospects of Recording. The Glenn Gould Reader. Ed. Page, Tim. New York: Vintage Books, 1984. 331-53.

Introduction

THIS INTRODUCTORY CHAPTER DEFINES THE PROBLEM SPACE OF THE THESIS; OUTLINES THE MULTIDISCIPLINARY RESEARCH BACKGROUND IN WHICH THE STUDIES PRESENTED IN THIS THESIS ARE SITUATED; DEFINES THE RESEARCH QUESTIONS AND CONTRIBUTIONS; AND SUMMARISES THE REMAINING SECTIONS OF THE THESIS.

1

1.1 INTRODUCTION

On rare occasions, people can experience extraordinary epiphanies in response to certain life events. These states of mind tend to happen when we experience a deep sense of exhilaration and enjoyment, when we are so involved in an activity that time seems frozen, and when concentration is so intense that self-consciousness disappears. The psychologist Mihaly Csikszentmihalyi spent decades investigating this state of mind, and eventually theorised the concept of *flow* or *optimal experience* (Csikszentmihalyi, 1991). Optimal experiences, he states, can be intentionally pursued, as they are more likely to happen in response to particular activities. Among these activities, music making is one of the most favourable to produce states of flow. This property is mostly attributable to the intrinsic rewards gained from undertaking and mastering this activity (O'Neill, 1999). Music making is, in fact, a particularly demanding activity involving nearly every cognitive function (Zatorre, 2005). Given adequate time and motivation, however, overcoming technical difficulties is feasible. What time and motivation alone cannot supply is the artist's sense and sensibility to transform ideas, emotions, meanings, and narratives into musical form.

In the last few years, multidisciplinary research has attempted to develop new computational methods to simplify music making in order to open this optimal experience to a broader audience. Researchers and practitioners from a number of different disciplines have been seeking to mitigate for the musician part of the complexity of music making. Novel computational systems have been proposed to support the composition of new musical pieces and provide simplified experiences of music performance. This research objective resonates with the strategies proposed by Serra, Bresin and Camurri (2007) when discussing the open issues in the Sound and Music Computing community. In their manuscript, the authors state that future research should "produce tools which can interact meaningfully with the user via sound, possibly integrated with other modalities. To do so, these tools will have to incorporate knowledge about sound perception and multimodal communication".

This thesis tackles the challenge of designing new musical systems for musically untrained listeners by research through design, a practice used to gain new knowledge via the act of making (Zimmermann, Forlizzi and Evenson, 2007). In particular, we started our exploration with the identification of an interactive metaphor that could mediate abstract musical meanings in concrete domains that everybody can easily understand. Interactive metaphors indeed allow "users to readily make inferences about how to operate unfamiliar user interfaces by mapping existing skills and knowledge from some familiar source domain" (Wilkie, Holland and Mulholland, 2013). To properly encode musical meanings for musically untrained listeners, the language on which the interaction metaphor relies has to meet a number of requirements. In particular, it has to (i) be available to everybody, (ii) be closely connected with music, (iii) "incorporate knowledge about sound perception and multimodal communication" (Serra et al, 2007), and (iv) be capable of dealing with music complexity. The language of emotions can be the best candidate as (i) most people are able to describe states of mind, (ii) music is often regarded as the language of emotions (Cooke, 1959), (iii) emotional perception in music have been deeply studied (Juslin & Sloboda, 2010), and (iv) the emotional space is extremely complex and distinctive (Russell, 1980).

This thesis proposes emotions as mediators of musical meanings in the context of interactive music making. The adequacy of this proposal is tested with the case study of The Music Room, an interactive installation designed to offer new experiences of music making to a wider audience (Morreale, Masu, De Angeli and Rota, 2013a; Morreale, De Angeli, Masu, Rota and Conci, 2014a). The Music Room interfaces players and Robin, an algorithmic composer that systematically converted user input described in emotional language into compositional rules, which are in turn used to direct the composition (Morreale, Masu and De Angeli, 2013c). This research through design led to new knowledge, which was formalised in MINUET, a design framework intended to stimulate creativity and reflections when designing musical interfaces.

1.2 THE MUSIC ROOM: AN INTERACTIVE INSTALLATION

The Music Room allows dyads of visitors to create an original classical-like musical composition by moving throughout a room. The control of the music played in the room is shared between Robin and the dyad, which can direct the emotional character of the music by varying their distance and speed. The distance between the members of the dyad determines the valence of the music (the closest the more positive) while their average speed determines the arousal (the fastest the more intense). The Music Room evolved through a two-year design process including a conceptual stage enriched by early evaluations of scenarios and storyboards and continuous testing of an evolving experience prototype.

The technical architecture of The Music Room is structured into two modules: a vision tracking system and Robin. A downward-looking bird's-eye camera mounted on the ceiling of the room detects the position and the speed of the couple via background-subtraction. A tracking algorithm processes this information, updating the couple's position and speed over time, and it supplies the extracted proxemic values to the system. Following the proposed mapping, the values are converted into the emotional cues of valence and arousal and are communicated to Robin, which reconfigures the generated music to match the intended emotion. The unpredictable nature of the algorithmic composition, combined with the two parameters of interaction (i.e. distance and speed), allows a broad range of original musical compositions and requires no musical expertise.

The evaluation primarily focused on analysing audience engagement, contributing with new insights to understand the experience of visitors in interactive art (Edmonds, 2010). Field evaluation was conducted during two live events opened to the general public. An integration of online observations, interviews, questionnaires, and offline analysis of log data and videos was performed. Results suggest that nearly all of the visitors experienced authentic enjoyment. Musically untrained visitors, in particular, referred to the experience as being remarkably creative. Audience engagement resulted from a number of very different behaviours. The most recurring one was dancing. Despite the fact that this behaviour did not come unexpected - as the synergy between music, movements and emotions is often associated with dancing - the motivations that disposed visitors to dancing remained ambiguous. Indeed, collected data failed to assess whether dancing was a reaction to music, as it normally happens when people dance, or they consciously moved to influence the music while dancing. The most controversial aspect was the involvement of the dyads in the music making. When prompted with this issue, only half of the participants reported that they felt in control of the music, while the other half declared that they were mainly following it. The analysis also failed to explain the reasons why the experience varied so dramatically among visitors, and to identify the main factors that accounted for the diversity of experiences.

A more controlled evaluation was organised to further investigate these open issues. With respect to previous exhibitions, the goal was to collect idiosyncratic interpretations rather than observing group reactions that would be expected to apply to everyone (Höök, Sengers and Andersson, 2003). With this end in view, 26 selected commentators, chosen on the basis of their professional profile and their artistic sensibilities, were invited to participate in and comment on the installation. In-depth interviews offered several interesting insights. In particular, results revealed that The Music Room offers a wide range of non-ordinary and engaging experiences, ranging from music making to dance, and from leisure to relaxation. The factors that most significantly contributed to defining visitors' engagement was the amount of information issued: the more information that was provided, the closer the experience was to that originally envisioned by the designer, to the detriment of the participants' creative engagement. Avoiding a clear narrative of use was associated with the highest potential for fostering a creative engagement, as opposed to fostering an engagement resulting from following pre-defined schemas.

1.3 ROBIN: AN ALGORITHMIC AFFECTIVE COMPOSER

The foundation of The Music Room is Robin, an algorithmic affective composer that automatically generates tonal music. The control over the composition is shared between the system and the visitors of the installation, who can determine its emotional character. To this end, different emotions are mapped to combinations of structural factors (i.e. what can be annotated in a score). Research in the psychology of music has identified a number of structural and performative parameters that contribute to shaping the emotional response of the listener (Gabrielsson & Lindström, 2010; Bresin & Friberg, 2011). Among structural factors, it is widely accepted that tempo and mode have the highest influence on emotional perception (for a review, please check Gabrielsson and Lindström, 2010). In most of the cases, perceived emotions are described as the result of the interrelation of two dimensions: *valence*, which refers to the positive vs. negative affective state, and *arousal*, which refers to rest vs. activation (Russell, 1980). Tempo is primarily responsible for determining the arousal of a musical piece, and it has a secondary effect on valence. Mode is responsible for determining the valence only.

Given its focus on a general population, one research question of this thesis investigates whether and how the emotional perception of tempo and mode varies across different levels of musical training. Related work provided controversial hypotheses about the influence of expertise in the emotional perception of music (Webster & Weir, 2005; Hargreaves & North, 2010; Castro & Lima, 2014), contrasting a vision of musical competence as an innate or a learned ability. However, until recently, there had been no systematic investigation of the separate effects of mode and tempo on this emotional dimension. To this end, an experimental study with 40 participants was conducted (Morreale, Masu, De Angeli and Fava, 2013b). Tempo (160BPM vs. 80BPM) and mode (major vs. minor) were manipulated in a 2*2 within-subjects design. Seven short piano pieces were ad-hoc composed and systematically manipulated using the four conditions, for a total of 28 snippets. For each snippet, participants were asked to self-report the perceived valence and arousal, and to indicate their liking. Results suggested that expertise has an impact on the emotional response to music but only on valence (defined by the combination of tempo and mode). In particular, a difference emerged in the evaluation of musical excerpts when tempo and mode conveyed diverging emotional information, suggesting that trained listeners are more sensitive to mode variations than non-musicians.

These results, coupled with findings from related work (mostly from Gabrielsson and Lindström, 2010), informed the design of Robin. Robin is the result of several years of research in algorithmic composition and the psychology of music (Morreale et al., 2013c). The compositional method is rule-based: the algorithm is taught a series of basic compositional rules of tonal music that are used to create original compositions in Western classical-like music with emotional character. At each new bar, a number of stochastic processes determine the best possible choice of tempo, mode, sound level, pitch register, pitch contour and rhythm. Users communicate to Robin in real time the emotions they want to express, defined in terms of valence and arousal, which immediately reconfigures the composition to produce matching music. For instance, when prompted to generate a happy melody (positive valence, high arousal), Robin sets the composition to high tempo, major mode, high sound level, ascending contour, and high register, and it enables theme repetitions. By contrast, a sad melody (negative valence, low arousal) triggers low tempo, minor mode, descending contour, low sound level, and low register

The validity of Robin in communicating different emotions in listeners was tested in an experiment using controlled conditions (N=33). Valence (positive vs. negative) and arousal (high vs. low) were manipulated in a 2*2 within-subjects design. Robin generated five snippets for each of the four conditions, for a total of 20 snippets. At the end of each snippet, listeners were asked to self-report the perceived valence and arousal, and to indicate their liking. Results showed that Robin correctly communicateed valence and arousal in nearly all cases. However, in cases of dierging conditions (high valence, low arousal and low valence, high arousal), valence received neutral values. The experiment validated the effective capability of Robin to communicate specific emotions in the listeners. This result, beyond supporting the adoption of Robin in interactive installations, entails another main implication. Robin showed its potential of being used in experimental studies aiming at testing the perception of structural factors. In addition to correctly eliciting the intended emotional response, the average liking of the stimuli generated by Robin was significantly higher than those used in the first experiment, which were composed by a human.

1.4 MINUET: A DESIGN FRAMEWORK

Visitors' feedback indicated that The Music Room greatly differs from any other traditional and digital musical instruments. In particular, the focus of the experience was transferred from the actual music making to the aesthetic, playful and emotional aspect of it. This change of focus suggested a reconsideration of the design process of musical interfaces. To date, design guidelines of musical interfaces have promoted a technology-oriented design. This approach well suits musical controllers that resemble the physicality and the objectives of acoustic instruments (Miranda & Wanderley, 2006). However, if we consider musical interfaces as an umbrella term that also includes interactive installations, this approach may prove unsuitable. Within interactive installations there are different characters, scopes and goals, and they usually focus on eliciting a particular user experience rather than a set of musical activities performed on an instrument. For instance, the Brain Opera aims to prompt reflection in the audience while participants actively operate on musical content (Machover, 1996). Similarly, Sonic Cradle helps visitors to achieve a meditative experience by controlling sounds through the exploration of their own respiration (Vidyarthi, Riecke and Gromala, 2012). Other installations try to foster active interaction between visitors. For instance, Mappe per Affetti Erranti reproduces the music in full-orchestration only when participants collaborate (Camurri et al., 2010). In a different design context, Piano Staircase[1] makes the activity of going up the stairs fun and social: an actual staircase placed next to an escalator is transformed into a giant piano keyboard and people can trigger relative notes by stepping the stairs.

To delve into the complexity of this design space we developed MINUET (Musical INterfaces for User Experience Tracking), a design framework intended to stimulate reflection when designing novel musical interfaces (Morreale et al., 2014b). MINUET offers a conceptual model for the understanding of the elements involved in the

1 http://www.thefuntheory.com/piano-staircase design of musical interfaces, providing a framework for this complex design space. In order to manage with the diversity of interactions, the design space is simplified to detecting patterns and suggesting insights that can assist designers' reflections. This challenge is addressed by clustering design elements from the point of view of the player's experience. Also, rather than providing a list of design metrics and heuristics, MINUET integrates a temporal dimension consisting of two sequential stages: analysing the goals of the interface and specifying how to achieve these goals.

The first stage of MINUET frames a conceptual model of the interface goal, in the form of a very high-level user story (Carroll, 2000). Designers are invited to inspect their goal through the lenses of *People*, *Activities* and *Contexts* (Benyon, Turner and Turner, 2005). *People* looks at the designer's objectives from the viewpoint of the targeted category of players and from the role of the audience. This point of view specifies the subjects *who* will engage with the interface (e.g. untrained musicians, sax players, music students). *Activities* questions *what* the envisioned interaction is, by framing the type of musical interface. It provides insights into the motivations of the players, analyses the relevance of music with respect to the player experience, and specifies the learning curve and possible collaboration among players. *Context* investigates the environment and the set-up of the interface, i.e. all of those elements that can assist in the identification of the interaction goals (*where/when*). The relevance of these entities varies according to the nature and the goals of the interface, and the priority scale has to be defined by the designers themselves.

The second stage of the design process involves designing the interaction in order to fulfil the designers' objectives. For instance, when designing for intuitive experiences, the number of interaction possibilities should be restricted, in order to guarantee easy access to the players. By contrast, musical controllers should enable players to manage a multitude of parameters, in order to have full control of the generated music. First and foremost, specifications must be considered according to the degree to which the player controls the music. Depending on the tasks and the targeted audience, interfaces can provide high- or low-level control of musical elements. In addition, input and feedback modalities contribute to determining the status of an interface. Furthermore, designers can suggest a particular strategy or interaction trajectory, providing complete documentation or including physical constraints embodied in the design of the interface. Conversely, they may seek to stimulate creativity, improvisation and adaptation: in this case, flexible and constraint-free interactions should be encouraged. The potential of MINUET for guiding designers throughout the design process of musical interfaces was tested with the design of The TwitterRadio, a tangible interactive installation designed to explore the social world of Twitter through music. The idea of The TwitterRadio is to browse a list of trending news and listen to the mood of public opinions on them. MINUET served to outline the design objectives and specifications of The TwitterRadio, framing the character of the installation, identifying design requirements and elaborating possible interaction trajectories.

1.5 CONCLUSION

In 1996, Tod Machover stated: "I now believe that the highest priority for the coming decade or two is to create musical experiences and environments that open doors of expression and creation to anyone, anywhere, anytime". This thesis takes up this challenge by focusing on the *anyone* part. A theoretically-grounded and empirically-validated framework for the design of new experiences of music making is presented. The first step consists of formalising a new interaction metaphor lying outside the musical domain that mediates the inherent complexity of music making. The language of emotion is proposed as the mediator, given the universality of the affective states and their natural connection with music. This switch opens up a number of research questions and exploratory studies, which are grounded on and contribute to different fields of research: musical interface design, algorithmic composition, and the psychology of emotion.

As regards research on musical interface design, this thesis contributes an in-depth analysis of the *suitability of emotions as an interaction metaphor* for musical installation, and of movements as an actualisation of this mediation with a specific focus on non-musicians. This analysis was supported by the design of The Music Room, an example of interactive installations based on the emotional metaphor. The Music Room was the result of a two-year iteration of design and evaluation cycles, which informed the operational definition of the concept of *engagement* and of new *evaluation protocols* based on the integration of evidences from different user-research techniques. Finally, the thesis introduces MINUET a *conceptual framework* to provide new perspectives for the design of musical interfaces.

As regards research on algorithmic composition, this thesis contributes Robin, a rule-based affective composer that automatically generates Western classical-like music in real time. As opposed to previous research, the quality of the music and its actual capability to communicate intended emotions in the listeners were tested with a formal experimental study. Furthermore, the potential of Robin to be used in interactive installations was exemplified with two case studies.

As regards research on the psychology of music, this thesis makes three contributions. Firstly, it contributes an experimental design to exploring listeners' perceptions of valence and arousal in response to divergent conditions of mode and tempo. Secondly, it provides new evidence to the on-going debate about the nature of musical competence investigating whether expertise influences the perception of emotions in listeners. Thirdly, it proposes the benefits of adopting an algorithmic composer as stimuli generator to be used in experimental studies aiming at testing the influence of structural factors.

The thesis report is structured in seven chapters. Chapter 2 defines the theoretical foundation of this work and analyses the current state of research in the areas related to the thesis topic, i.e. musical interface design, psychology of music, and algorithmic composition. Chapter 3 reports an experimental study aimed at understanding the ability to recognise emotion in music; presents Robin, the algorithmic affective composer; and it reports an experimental study aimed at testing the validity of Robin in communicating different emotions in listeners. Chapter 4 presents the conceptual design and the technical implementation of The Music Room. Chapter 5 presents the evaluation of visitors' experience with The Music Room, proposing new insights for understanding audience engagement with interactive art. Chapter 6 proposes MINUET, a design framework for the understanding of the elements involved in the design of novel musical interfaces. Chapter 7 concludes with a discussion of the thesis findings and the implications for the research areas of interest.

Related Work

THE THEORETICAL AND EMPIRICAL INVESTIGATIONS OF THIS THESIS ARE EMBEDDED IN SEVERAL RESEARCH AREAS. SPECIFICALLY, THIS CHAPTER REVIEWS LITERATURE FROM INTERACTION DESIGN, THE PSYCHOLOGY OF MUSIC, AND ALGORITHMIC COMPOSITION.

2

2.1 INTRODUCTION

This chapter explores the multidisciplinary foundation that underlies the studies discussed in this thesis. Interactive music making is a rapidly growing multidisciplinary area that relates very closely to interaction design, psychology of music and algorithmic techniques.

The objective of this thesis is to design new experiences of music making based on novel, intuitive interaction metaphors. This theoretical framing is tested in The Music Room, an interactive installation that allows visitors to direct the emotional character of an algorithmically generated music. The design process is partially influenced by that of similar installations, and draws inspiration from the design heuristics of musical interfaces, which are surveyed in Section 2.2. This survey revealed some gaps in existing design frameworks: in particular, interactive installations are not adequately considered. This gap led to the development of MINUET (Chapter 6), a design framework for musical interfaces, that helped to position The Music Room in a precise design configuration, and assisted in guiding the evaluation. To evaluate audience experience previous research discussing engagement with interactive installations was considered.

The proposed interaction metaphor suggests exploiting emotions to mediate musical complexity. In order to create music with emotional character, the associations between alterations to musical factors and changes in perceived emotion need to be investigated. To explore this issue, a research in the psychology of music was conducted and is reported in Section 2.3. Findings from these studies informed the design of Robin, an algorithmic composer that adjusts composed music to users' desired emotions in real time (Chapter 3). The compositional strategies driving the stochastic processes that create harmony, melody and rhythm draw inspiration from existing approaches to algorithmic composition (Section 2.4).

2.2 INTERACTIVE MUSIC MAKING

Neuroscientists have long been fascinated by the unique demands made on the nervous system when making music (Schlaug, Norton, Overy and Winner, 2005; Zatorre, Chen and Penhune, 2007; Patel, 2010; Koelsch, 2011). Despite not being strictly necessary to the survival of human species, music is in fact one of the most complex human activities. Robert Moog, one of the pioneers of electronic music, estimated that a skilled musician is able to generate about 1,000 bits/sec of meaningful information[1]. Music making is traditionally regarded as a two-stage art consisting of composition and performance (Brauneis, 2014).

1 *Moog, R. Keynote speech at NIME 2004.*

Composition produces a score, i.e. a "stable, visually perceptible representation of melody, harmony and rhythm, using a system of mostly discrete notation". A score is then realized in performance, i.e. "a real-time, low-deliberation, no-editing activity that is evanescent, unrepeatable, purely aural, and continuous" (Brauneis, 2014). Mastering both the knowledge of music theory that is necessary to compose quality music, and the practice of instrument playing that is necessary to perform correctly, requires years of study and dedication. On the one hand, the learning effort necessary to master composition and performance provides enjoyment for setting and following longer-term goals (Csikszentmihalyi, 1991). On the other hand, it limits most people to enjoying music only by listening or dancing to melodies created by somebody else. Only trained musicians can master the intrinsic difficulties of composing and performing quality music.

Music composition, in most genres, is grounded in complex rules that describe the progression of melody, harmony and rhythm. The compositional rules, which act as a reference and support to composers in their pieces, require time and practice to be memorised and applied. At the present time, novel computational systems have been developed to mimic the composer's ability to write music by applying these rules via algorithmic techniques (Cope, 2005; Nierhaus, 2009; Boenn, Brain, De Vos, Ffitch, 2011). In these systems, compositional rules, manually or statistically trained, drive a number of stochastic processes that, in turn, produce original scores. Other computational systems (Suggester[2], Guitar Pro[3], Fiddlewax Pro[4]) also offer support to composers by facilitating song composition and arrangement. These systems propose to replace or complement theory books by providing easily accessible knowledge about music theory in an interactive way.

Music performance is a highly demanding cognitive and motor challenge that demands skill development and automation. Performing even a simple musical piece requires a host of skills to control pitch and rhythm over an extended period of time (Zatorre et al., 2007). To do so, a performer needs to process complex musical notation, translate it into bimanual motor activity and memorise intricate musical passages (Schlaug et al., 2005). At the present time, novel Digital Musical Instruments (DMIs) have been aiming at easing this process by replacing or complementing traditional instruments. GarageBand for iPad[5], for instance, offers simplified virtual models of traditional instruments (e.g. piano, string instruments, guitars), to ease performers' cognitive and motor efforts (Figure 2.2). The featured Smart Instruments allows users to create music by performing operations on auto-generated grooves and riffs. In the view of Ruismäki, Juvonen and Lehtonen (2013), the quality of the music generated by GarageBand is adequate, and the experience is fun. Other DMIs seek to ease music performance, discarding analogies with traditional instruments and redesigning the interaction with musical contents from scratch. The Reactable is one of the most successful attempts to provide novel solutions to create electronic music (Jorda, Geiger, Alonso and Kaltenbrunner, 2007). Music is directed by placing and moving tangible blocks that are associated with a particular acoustic or musical element, on a touchscreen tabletop.

2 https://itunes.apple.com/us/app/suggester-chord-progression/id504740787

3 http://www.guitar-pro.com/en/index.php

4 https://fiddlewax.com

5 https://itunes.apple.com/us/app/garageband/id408709785

Figure 2.1. Curtis Bahn playing the S-Bass vs. a Reactable performance.

Figure 2.2. Screenshot of the Smart Instrument feature of GarageBand for iPad.

A simple design space of interfaces for music making can be described along a continuum that stretches from Digital Musical Instruments (DMIs) to interactive installations (Wanderley & Orio, 2002; Birnbaum, Fiebrink, Malloch and Wanderley, 2006). DMIs are technological artefacts that reproduce traditional instruments; interactive installations are artistic exhibits that allow music making with hardware that bears little or no resemblance to instruments. More articulated frameworks have focused mainly on DMIs by differentiating, for example, among augmented traditional instruments, instrument-like controllers, instrument-inspired controllers, and alternate controllers (Miranda & Wanderley, 2006). This categorisation is device-oriented: while considering the similarity to traditional instruments and the technology featured in the controller, it fails to account for the profound variations affecting the experience of the player. As an example, consider the difference between a bass player using an augmented bass like Curtis Bahn's S-Bass (Bahn & Trueman, 2001) and a Reactable player (Jordà et al., 2007) (Figure 2.1).
In the other pole of this design space reside interactive installations, i.e. those systems "that are only realized through a participant's actions, interpreted through computer software or electronics, and those actions do not require special training or talent to perform" (Winkler, 2000). Among interactive installations, there are very different characters, scopes and goals, and they usually emphasise the user experience rather than a set of musical activities performed on an instrument. Section 2.2.1 surveys previous studies that attempted to define the design space of DMIs and interactive installations. Section 2.2.2 narrows the investigation to existing interactive musical installations. Research on the evaluation of interactive musical installations is very limited; thus, the investigation was extended to the broader design area of interactive art. Specifically, we examined methodologies to understand (Section 2.2.3) and to evaluate (Section 2.2.4) visitors' engagement with interactive artworks.

2.2.1 Design space

Over the last few years, the investigation of the design space of interfaces for music making has been arousing an increasing interest. Most of the research in this area has restricted the focus to DMIs, disregarding interactive installations. However, the findings discussed in some of these studies (Johnston et al., 2008; Wallis et al., 2013) can be partially extended to the broader design space that also includes interactive installations.

Two meta-reviews suggested how to catalogue DMIs from different perspectives. Drummond (2009) analysed the different approaches to the definition, classification and modelling of DMIs, while Marquez-Borbon and colleagues (2011) surveyed the methodological approaches from a design perspective. Johnston, Candy and Edmonds (2008) addressed this challenge from the general perspective of the interaction between players and musical instruments. Thus, they differentiated between instrumental (the musician has control over every aspect of the instrument), ornamental (the system has control), and conversational (shared control). Similarly, Jordà (2004) centred the framework on the relation between player and instrument, discussing issues of balance (between complexity and simplicity), playability, learning curve, and instrument efficiency. A recent study presented by Wallis, Ingalls, Campana and Goodman (2013) narrowed the analysis to the musical instruments that inspire long-term engagement, and proposed seven heuristics to describe their qualities: incrementality, complexity, immediacy, ownership, operational freedom, demonstrability, and cooperation.

Other frameworks classified DMIs based on the type of gesture employed to control the musical interface. Overholt (2009), for example, distinguished between intuitiveness and perceptibility of gestures, and between ergotic (gestures used to manipulate physical objects) and semiotic (gestures used to communicate meaningful information, such as thumb up) gestures. From a different perspective, Hunt and Kirk (2000) analysed the strategies for mapping gestures onto synthesis parameters differentiating between analytical and holistic cognitive modes. The analytical mode is directed towards a particular goal, such as following a score, or mapping a sound into an instrument. Conversely, in the holistic mode, the listeners perceive the overall effect of the music, disregarding individual instrumental voices.

Only a few studies have broadened the scope of the design space of musical interfaces by additionally considering interactive installations (Wanderley & Orio, 2002; Blaine & Fels, 2003; Birnbaum et al., 2006; Erkut, Jylhä and Disçioglu, 2011). Among these, Birnbaum and colleagues (2006) based this broader design space on seven dimensions: required expertise, musical control, feedback modalities, degrees of freedom, inter-actors, distribution in space, and role of sound. Blaine and Fels (2003) also proposed a number of design dimensions with a focus on collaborative systems: physical devices, type of interaction, learning curve, pathway to expertise, level of physicality between players, directed interaction and musical range.

Other studies have attempted to address the issue of evaluating control and usability of DMIs and interactive installations. Among these, Wanderley and Orio (2002) suggested drawing inspiration from HCI techniques for the evaluation of the control and usability of musical interfaces. This framework is based on learnability, explorability, feature controllability, and timing controllability. Erkut and colleagues (2011) addressed this issue from the viewpoint of the modality of interaction: the interface can employ a simple modality (visual or auditory), or multiple modalities by integrating simple modalities. It is worth noting how most of the dimensions presented in this last set of studies greatly differ from those traditionally adopted when classifying DMIs only (e.g. role of sound, distribution in space, feedback modalities and type of interaction).

The contribution of these works towards a better understanding of the design space of interfaces for music making is indisputable. However, the actual application of any such framework in the design of novel interfaces is nearly nonexisting. A possible reason is that the quality of a framework should be based on how it helps designers to stimulate creativity and allows them to delve into the design process. Furthermore, the increasing number of interactive installations operating in the musical domain suggests that further research should be conducted in this area.

2.2.2 Collaborative music making

Interest in research on the design of DMIs and interactive installations for music making has been growing in the last decade (Blaine & Fels, 2003). Most of them rely on collaborative behaviour between groups of visitors. Following the emphasis of this thesis, we cluster this research into two categories: i) musical interfaces that target an audience that have at least some musical training; ii) musical interfaces that target untrained users.

The first category counts a number of tangible interfaces, such as the Reactable (Jordà et al., 2007). Similarly, Jam-O-Drum exploits tabletop technology to foster collaborative improvisations (Blaine & Perkis, 2000), and the AudioPad allows performers to control sound synthesis via tangible interaction (Patten, Recht and Ishii, 2002). These systems combine visually-pleasing aesthetic with collaboration-oriented features. In this framework, rather than operating on low-level musical parameters, as with traditional instruments, players encode their musical meanings in higher-level musical structures (e.g. sequencers, scale selection, envelopes, beat) that decrease the cognitive and motor effort required for making music. Despite of this, higher-level musical structures are still unlikely to convey any meaning to untrained users.

The second category counts a novel wave of collaborative systems that specifically address untrained users. These systems exploit the concept of active listening (Rowe, 1992): users can interactively control the musical content by modifying it in real time while they are listening to it (Camurri, Canepa and Volpe, 2007). Following this concept, several works have sought to enable people to shape musical content through collaborative interaction (Machover, 1996; Camurri, Volpe, Poli and Leman, 2005; Camurri, Varni and Volpi, 2010). In the following paragraphs, we describe a few of these projects.

In Mappe per Affetti Erranti a group of people can experience active listening by exploring pre-composed music and navigating a physical and emotional space (Camurri et al., 2010). The installation specifically encourages collaboration, as music can only be appreciated in its full complexity if the participants cooperate with each other by moving throughout the space. The interaction space is divided into several areas, each associated with a melody, which is triggered only when a visitor is present in the area. Additionally, different expressive intentions can be selected by performing expressive gestures. For instance, hesitant behaviours cause the corresponding melody to be produced at the volume of a whisper. Mappe per Affetti Erranti encourages strangers to collaborate in the artistic creation of music. This collaborative approach was also employed in TouchMeDare (van Boerdonk, Tieben, Klooster and van den Hoven, 2009). Two or more people can make music by interacting through a canvas, and pre-composed music samples are only triggered when the canvas is simultaneously touched by more than one user.

Other studies have mediated collaboration between participants using mobile phones. Among them, Sync'n'Move allows users to experience music with social interaction (Varni, Mancini, Volpe and Camurri, 2010). Two users freely move their mobiles, and the complexity of music orchestration adjusts proportionally to the synchronisation of their movements. Accordingly, if synchronisation fails, there is no orchestration at all; if synchronisation is only partially achieved, the orchestration is quite elementary; in the case of perfect synchronisation, the orchestration is complete. Based on a similar design, MoodifierLive allows the control of automatic music performances through a number of interaction modes based on user collaboration via gestures performed using a mobile phone gestures (Fabiani, Dubus, and Bresin, 2011).

Other works have endeavoured to exploit human expressiveness and emotions to influence the status of the system (Camurri et al., 2005; Camurri et al., 2010), using body gestures (Mancini, Castellano, Peters and McOwan, 2011) and dance movements (Camurri, Lagerlöf and Volpe, 2003). Most these systems exploit video analysis techniques ranging from simple position tracking (Camurri et al., 2010) to complex full-body and hand-movement tracking. One of the most elaborate systems is EyesWeb, a platform for the design and development of real-time multimodal systems specifically designed to track the gestures of performers and to extract expressive content (Camurri et al., 2000).

The interaction metaphors proposed by the surveyed works are innovative and have contributed in different ways to empower untrained users to actively operate on musical elements. However, in most of the studies, the design process of the interfaces is only partially described, hampering the establishment of best practices and neglecting insights for other researchers interested in building similar interfaces. Furthermore, there is very little if any evaluation of the player's experience. In some exceptional cases, the evaluation is limited to administering questionnaires (van Boerdonk et al., 2009; Varni et al., 2010) or to testing the quality of the system (Fabiani et al., 2011). To remedy this gap, we examined the broader design space of interactive art in order to review methods to understand visitors' experience.

2.2.3 EVALUATING EXPERIENCE IN INTERACTIVE ART

The interest of the scientific community in interactive (art) installations is slowly but steadily growing (England, 2012). This accounts for the increasing number of collaborations between researchers, practitioners and artists. Yet, specific challenges related to understanding audience experience and engagement remain unsolved (Edmonds, 2010). In the last few years, designers and artists have come to recognise the need to evaluate experience with interactive installations (Edmonds, 2010) using a diversity of `methods. Candy, Amitani and Bilda (2006) described the research methods most often adopted to study the interactive art experience: direct observation, observation via video recording, contextual interviews, and video-cued recall.

Following an ethnography-inspired approach, Morrison and colleagues (2007) proposed a method for evaluating the human experience in relation to interactive art: according to them, the artist should become a hidden participant, take part in informal discussions with participants in situ, and ask them to complete formal questionnaires. Following an expert-based approach, Amabile (1996) proposed the consensual assessment technique to measure the creativity of products and processes, employing the subjective assessment of experts.

Šimbelis and his colleagues (2014) proposed a method based on a variant of Gavers' cultural commentators to facilitate a dialogue between Metaphone, an artwork aiming at conceptualizing machine aesthetics (i.e. an art style exposing the inner aesthetics of the technology), and a group of artists. In the original proposal by Gaver (2007), cultural commentators were selected outside of the native community of practice "as resources for multi-layered assessment" of pieces of design. Invited commentators were asked to reflect upon the work by producing interpretations in a form that mirrored their professional backgrounds (e.g. a documentary for filmmakers, a storyboard for writers). With respect to the original method, this approach was applied to the evaluation of interactive installations, but with two changes: the commentators were selected from inside the community of interest, and their interpretations were collected in the form of verbal accounts. Despite these alterations, this method allowed the collection of interesting feedbacks and critiques, confirming that commentators had experienced the installation as originally intended by the designer.

2.2.4 ENGAGEMENT IN INTERACTIVE ART

Edmonds, Muller and Connell (2006) discussed engagement with interactive art in terms of three attributes distributed along a temporal dimension: attractors, sustainers and relaters. Attractors are those features that initially catch visitors' attention and cause passers-by to notice the work. Sustainers are those peculiar features of the work that, once the audience has been attracted, keep the audience engaged. Relaters are those features that ensure a long-term interest in the work once the initial pleasure has worn off, stimulating the audience to return to the work on future occasions.

Engagement with interactive installations is believed by Morrison, Mitchell and Brereton (2007) to vary widely among people. It would be inappropriate to expect the generation of uniform experiences and to consider an artwork successful only if every participant has engaged with it as expected by the designer. Design should in fact anticipate multiple interpretations of the system (Sengers & Gaver, 2006) given by the "user's interpretation, understanding, attitudes, personality and expectations of computer culture" (Hook et al., 2003).

Elaborating on the role of subjective variability, Bilda, Edmonds and Candy (2008) highlighted the importance of subverting expectations to foster engagement with interactive art. In their view, engagement is a transformative dialogue between the participant and the artwork. Participants explore installations based on their expectations: if expectations are not met, they may become curious, continue exploring the system and eventually reconsider their intentions and expectations. To this end, making the intended result ambiguous to the audience can foster their creative engagement: an uncertain mode of interaction has the highest potential for creative engagement, as users can derive creative outcomes or increased understanding.

This position is shared by Gaver and colleagues (2004), who claimed that designing ambiguous applications, in terms of their interpretation and meaning, involves avoiding a clear narrative of use. Unclear narratives of use reduce the worry of failure, as the outcome of the interaction cannot be formally assessed (Csikszentmihalyi, 1991). However, the unpredictability of a system should be balanced by predictability. Predictable interactions risk boring their audience and lessening audience interpretation, but high unpredictability risks alienating the audience “making them feel stupid and out of control entirely” (Höök et al., 2003). On a similar note, Kwastek (2013) claimed that the relationship between chance and control is central to the experience in interactive art: exploring the functionality of the work is an important component of audience engagement. These concepts originate from HCI research on designing for appropriation in interactive systems. According to Alan Dix (2007), appropriation occurs when “people do not ‘play to the rules’: they adapt and adopt the technology around them in ways the designers never envisaged”.

2.2.5 THESIS CONTRIBUTION

The literature review on existing interactive musical installations revealed a research gap on the design and evaluation process of these systems. We argue that in such a design-oriented field of research, prototyping and experience evaluation should not be overlooked. Endorsing this belief, this thesis aims to establish new knowledge and insights for the design and evaluation of musical interfaces. In particular, it provides three main contributions:

1. Chapter 4 presents the design of The Music Room, an interactive installation to open new experiences of music making to a wide audience. The contribution of this work to the design community is the adoption of a new interaction metaphor based on emotions to mediate the complexity of music making. Furthermore, we describe and discuss a process for the conceptual design based of interactive installations based on different prototyping techniques.

2. Chapter 5 proposes a method to support the evaluation of the experience of visitors with interactive musical installations. In particular, we propose the integration of evidence from quantitative and qualitative methods to understand audience experience and engagement. Results also inform the concept of engagement with interactive art.

3. Chapter 6 presents MINUET, a framework intended to (i) stimulate reflection when designing novel musical interfaces, and (ii) help designers to position, shape and evaluate their systems. As well as incorporating all of the varieties of musical interfaces, MINUET shifts the focus of previous frameworks, defining a design space of musical interfaces centred on players’ experience.

2.3 MUSIC AND EMOTION

Researchers and musicians have long been challenged by the task of understanding the factors that contribute to imparting a specific emotional connotation to music (Hevner, 1935). Categorising and measuring emotions are demanding tasks that become even more challenging when the emotions are associated with music. The expressive strategies adopted by composers and performers are, in fact, not always accessible to conscious introspection (Juslin & Timmers, 2010). The research that has explored this issue provides evidence for the criticality of this challenge. Over 100 experimental studies have sought to map musical factors to emotional expressions. For a complete review, interested readers can refer to (Juslin & Sloboda, 2010).

Effectively, research on the psychology of music suggests that the interpretation of the emotional expression in music depends both on structural and performative rules (Livingstone, Mühlberger, Brown and Loch, 2007). Structural rules are selected by the composer and relate to the music score itself. Performative rules relate to the performer's expressive interpretation of the score. Each rule prescribes the association between musical factors and the emotional effect produced by their modification. Structural rules are defined by combining structural factors such as tempo, mode, sound level, pitch range and pitch contour (Temperley, 2004; Ilie & Thompson, 2006; Gabriellson & Lindström, 2010). Performative factors are defined by combining performance factors such as articulation, timing, phrasing, register, timbre, and attack (Bresin & Friberg, 2000; Juslin & Sloboda, 2010; Gabriellson & Lindström, 2010; Bresing & Friberg, 2011). Different combinations of these factors offer a wide range of variability in the spectrum of emotional communication. For instance, in the last movement of the Symphony No. 6 Pathetique, Tchaikovsky expressed his deep sorrow through use of the minor mode and slow andante lamentoso tempo; in contrast, Vivaldi evoked happiness with the major mode and fast allegro tempo in his Primavera from Le Quattro Stagioni.

The experimental study of the human emotional response to music is complex and directly relies on the theoretical background on emotion, as well as the experimenter's ability to manipulate music and measure the listener response. Section 2.3.1 discusses different models of emotions, focusing on their adequacy in accounting for the emotional response to music.

2.3.1 MODELS OF EMOTIONS

The first studies that systematically probed the nature of emotions used a categorical approach. This approach postulates that all emotions can be derived from a finite number of monopolar factors of universal and innate basic affects. Usually, these include happiness,

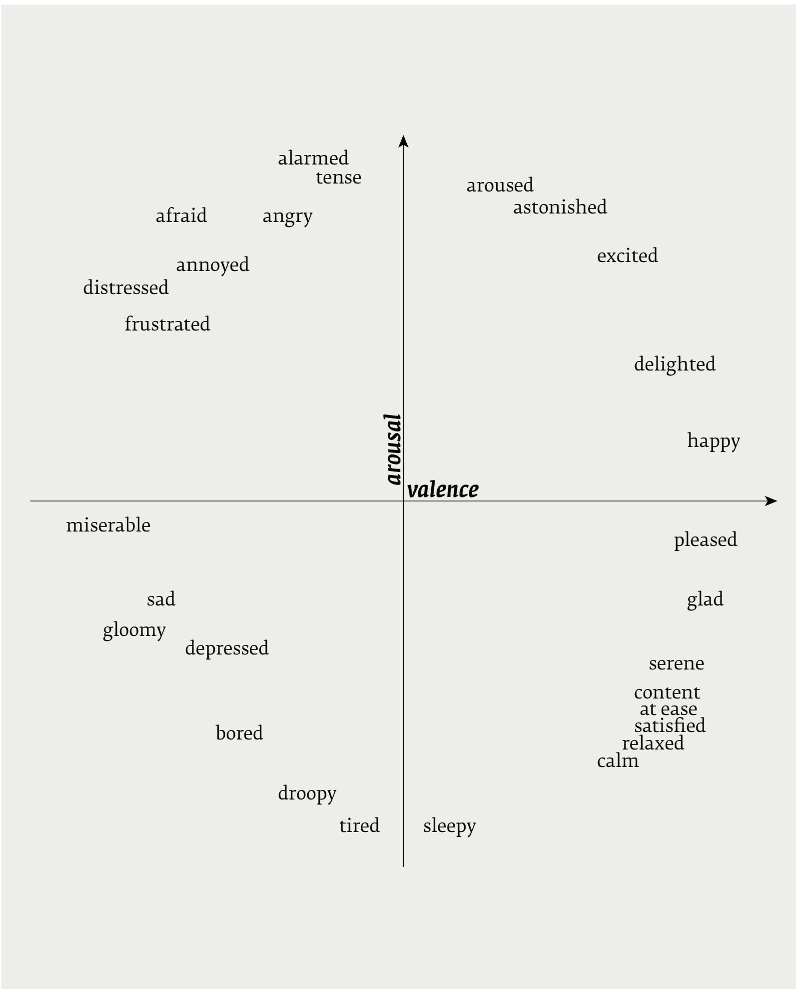


Figure 2.3. Russell's circumflex model of emotions (image from Russell, 1980).

fear, anger, disgust and sadness (Ekman, 1992). Despite being employed by several experimental studies in psychological, physiological and neurological studies, the main limitation of this approach is related to disagreement about the actual number and labels of categories (Zentner & Eerola, 2010). Furthermore, research on the psychology of music suggests that this theoretical background can introduce a major confound in the evaluation of emotion in music. Indeed, if the identification of the listener's emotion is mediated by verbal expression, it will be affected by his or her own ability to understand verbal labels (Gagnon and Perez, 2003).

The first arguments supporting the claim that affective states are not independent but are systematically related to each other dates back to 1952, when psychologist Harold Schlosberg (1952) derived a circular representation of emotions described along the dimensions of pleasantness-unpleasantness and attention-rejection. Two years later he extended the model with a third dimension, sleep-tension (Schlosberg, 1954). In the following years, other two- and three-dimensional models of emotions were proposed. The two-dimensional model was supported by Russell (1980), who argued that additional dimensions only account for a small proportion of variance, and that there is minimal consensus about their interpretation. In the same paper, Russell presented his renowned circumflex model of affect (Russell, 1980).

In this model, emotions are described as a continuum along two dimensions: valence, which refers to the pleasure vs. displeasure affective state, and arousal, which refers to the arousal vs. sleep difference (Figure 2.3). Despite being widely adopted in several fields of research, the limitations of this model were acknowledged by the author himself (Russell, 1980). Among others shortcomings, he noted that the affective states in which the two dimensions are convergent (i.e. positive valence and high arousal, and negative valence and low arousal) occur more frequently than do the affective states in which they diverge (Russell, 1980).

For the purposes of classifying emotions in music, both categorical and dimensional approaches have been widely employed (Zentner & Eerola, 2010; Bresin & Friberg, 2011). For example, the categorical approach was adopted by Gabrielsson (1995) when studying the set of basic emotions that can be elicited in listeners. He identified anger, sadness, happiness, fear, solemnity and tenderness. In some cases, musically inappropriate categories, such as disgust and surprise, have been replaced with more fitting emotional categories such as tenderness and peacefulness (Gabrielsson & Juslin, 1996; Zentner & Eerola, 2010).

The dimensional approach has been adopted by dozens of studies in the psychology of music (Schubert, 1999; Ilie & Thompson, 2006; Juslin & Sloboda, 2010), often employing the valence and arousal dimensions (Russell, 1980). This approach is considered to be better able to determine gradients of emotions more effectively than fixed categories can, although a general consensus posits that the two-dimensional model may not be able to account for all of the variance in music-mediated emotions (Ilie & Thompson, 2006; Zentner & Erola, 2010). For instance, this model closely locates emotions that are commonly regarded as distant, like anger and fear, which are both negatively valenced and highly active (Zentner & Erola, 2010). Despite the lack of agreement concerning what kind of model of emotions provides the best fit for perceived emotions in music (Zentner & Eerola, 2010), the literature reflects a strong preference for the dimensional approach, as it allows a finer assessment of emotions and it has better semantic resolution (Schubert, 1999). In this thesis, we employ the dimensional approach to systematically manipulate emotional connotation based on structural factors.

2.3.2 STIMULI SELECTION

Technical limitations have long restricted the possibilities for researchers to conduct experimental studies in the field of the psychology of music. The very first experiments on music perception actually involved live performances by professional musicians (Gilman, 1891; Downey, 1897). Listeners reported a variety of emotions, and some tentative relationships between musical factors and emotions were proposed (for instance, descending triads were perceived by listeners as sad). Subsequently, modern recording and synthesis techniques have allowed experimenters to gain partial control of stimuli either by using existing tracks or by playing pre-composed musical sequences. Using existing music ensures good ecological validity (i.e. musically acceptable pieces), but conclusions regarding the effects of individual musical parameters remain merely tentative (Robazza, Macaluso and D'Urso, 1994; Vieillard, Madurell, Marozeau and Dacquet, 2005; Gabrielsson & Lindström, 2010).

An alternative approach is to systematically manipulate separate factors in short ad hoc composed sound sequences. If, on the one hand, this approach reduces the ecological validity of the music, it improves the experimenter's control, permitting the manipulation of separate factors, and thus allowing the researcher to systematically analyse the effects of the tested factors (Gabrielsson & Lindström, 2010). In some extreme cases, a number of musical factors (e.g. intervals, mode, tempo, tone) were manipulated and tested in short sound sequences without musical context. This method, however, impoverishes the ecological validity. A trade-off between these two approaches combines ecological validity with a systematic manipulation of musical factors within a musical context (existing pieces of music or ad hoc composed excerpts). Still, manipulating some of the factors in existing music might result in musically unnatural stimuli (Gabrielsson & Lindström, 2010).

Gabrielsson and Lindström (2010) claimed that there is not a single correct alternative to select musical stimuli in a listening experiment, as all of them are affected by important limitations. However, when testing the influence of musical factors on the elicited emotions, there is a need to combine ecologically valid and musically natural music with systematic control over the tested factors.

2.3.3 Measuring responses

Measuring listeners' emotional responses to stimuli is another demarcation line in music perception studies. This task is particularly complex, as emotions are subjective phenomena that differ among individuals, contexts and moments. Fortunately, research in psychology provides several techniques to assess these subjective emotional responses with some reliability. For those interested in a comprehensive review of this topic, we refer the reader to (Zentner & Eerola, 2010).

Most studies in the psychology of music have measured perceived emotions, following a self-reported approach (Zentner & Eerola, 2010). In this approach, participants subjectively assess their experienced emotions by means of questionnaires, and at times using pictorial versions (Bradley & Lang, 1994). The emotional scales adopted to retrieve this information are based on categorical or dimensional models of emotions, with a strong preference for the circumflex model (Zentner & Eerola, 2010). Self-reported emotions can be collected either post-performance or continuously. Most studies have adopted the former approach. In this case, a single retrospective rating is provided after stimulus exposure (Zentner & Eerola, 2010). In other cases, the perceived emotion is continuously recorded while listening to music (Schubert, 2010). This alternative allows the detection of periodic fluctuations in response to musical variations (Schubert, 2001), but it exerts high cognitive loads on the participants (Zentner & Eerola, 2010). Another limitation of this approach refers to the participants' bias, as listeners can sometimes guess the objective of the research, and may accordingly adjust their rating to conform to perceived expectations (Västfjäll, 2010).

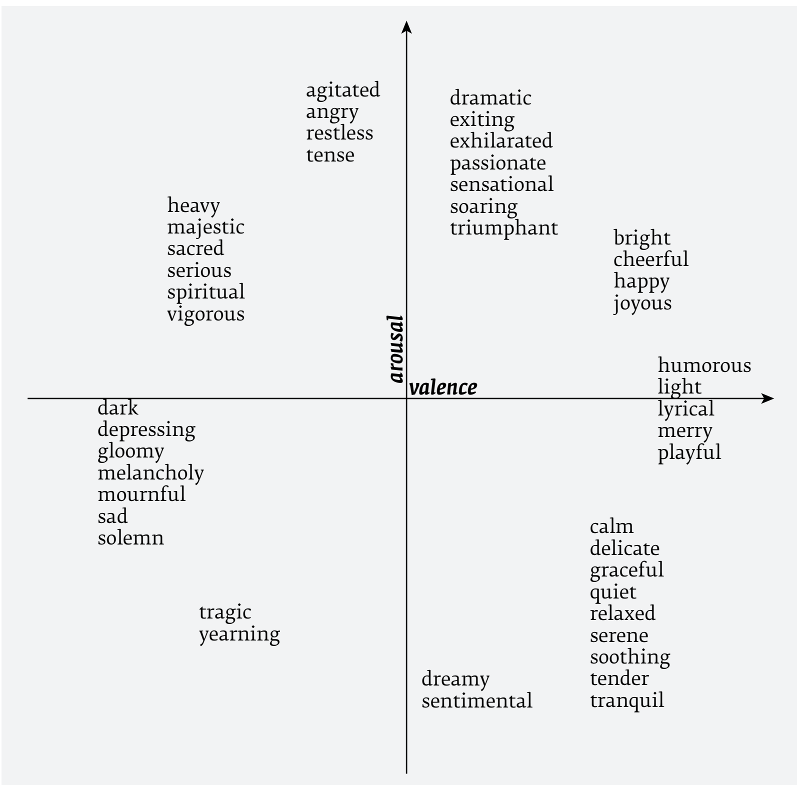


Figure 2.4. Hevner's circular configuration of emotions with updated terminology from Schubert, 2003 - which was readapted using the circumflex model.

To overcome these limitations, in the last few years, a few studies of emotions in music have applied measures of affective states using psycho-physiological sensing (Hodges, 2010; Trost, Ethofer, Zentner and Vuilleumier, 2011). This method can capture emotional aspects that might not be plainly evident to listeners. However, this method is still in its infancy and may require the listener to wear intrusive equipment. Most research continues to employ self-reported measures (Västfjäll, 2010). In the next subsection, we summarise, the most relevant findings obtained through this approach.

2.3.4 THE EFFECT OF MODE AND TEMPO

One of the first studies aiming to identify the association between musical factors and changes in emotional expression was performed by Kate Hevren in 1937. First, she arranged a large number of emotions in eight clusters placed in a circular configuration. Then, each cluster was associated with an adjective, and she investigated the effect of musical factors on each cluster. This adjective clock, which clearly resembles Russell's circumflex model (Russell, 1980), was readapted in the dimensions of valence and arousal by Schubert (2003), who redistributed the emotions into nine clusters (Figure 2.4).

In recent years, research on the psychology of music has showed renewed interest in understanding the musical factors affecting emotion (Bresin & Friberg, 2000; Meyer, 2008; Fritz et al., 2009; Gabrielsson & Lindström, 2010; Bresin & Friberg, 2011). A comprehensive guide reporting the main findings can be found in the book by Juslin and Sloboda (2010), which describes the emotional response to both structural and performative factors. In this thesis, we concentrate mainly on structural factors (i.e. those related to the musical score itself), as our objective is to algorithmically generate the musical structure. In particular, our review focuses on tempo[1] and mode, generally recognised as the most expressive structural elements, with a subtle predominance of tempo (Gundalach, 1935; Rigg, 1964; Juslin, 1997; Gagnon & Perez, 2003; Gomez & Danuser, 2007). Most of the related work adopted the circumflex model to classify listeners' emotional responses to the alteration of structural parameters. Exploiting the dimensions proposed by this model, tempo was proved to have a major impact on arousal and a minor impact on valence, while mode only impacts valence (Gagnon &

6 Tempo can be considered both as a performative and a structural factors. When referred as a strucutral factor, it is better described as "note density". Aware of this distinction, for the sake of simplicity, in the remaining of the thesis, we will simply refer to "tempo".

Perez, 2003). With respect to tempo, high arousal is associated with fast tempo and low arousal with slow tempo. With respect to mode, positive emotions are associated with major mode and negative emotions with minor mode. Also, but to a lesser extent, fast tempo stimulates positive emotions and slow tempo stimulates negative emotions. These findings are summarised in Table 2.1.

Table 2.1. Influence of mode and tempo in valence and arousal.

		VALENCE	AROUSAL
MODE	major	++	
	minor	--	
TEMPO	fast	+	++
	slow	-	--

Still debated is the emotional response when the combination of tempo and mode diverges (i.e. music played with major mode and slow tempo, or minor mode and fast tempo). These conditions should have opposite effects on perceived valence. Gagnon & Perez (2003), testing this hypothesis with 32 untrained musicians, confirmed that mode and tempo have indeed the highest relevance for communicating emotions, and they showed that this result is stronger in cases of convergent conditions (i.e. major mode combined with fast tempo or minor mode combined with slow tempo). When the two conditions are divergent, participants seemed to rely more on tempo than mode. Webster & Weir (2005) investigated this issue, taking into consideration listeners' expertise. Their result disagreed with Gagnon and Perez (2003), showing that in the case of diverging conditions, listeners reported similar, neutral values. A limitation of both studies was their reliance on dichotomous classification of emotion (happy vs. sad).

This subsection suggests that, to some degree, the effect of mode and tempo on valence and arousal have already been investigated. However, more research is required to better understand emotional response in the case of divergent conditions of mode and tempo. In particular, given the double action of mode and tempo, the perception of valence when mode and tempo suggest opposite emotions remains unclear.

2.3.5 THE INFLUENCE OF OTHER STRUCTURAL FACTORS

In addition to tempo and mode, other structural factors have been found to have an influence on the perceived expressiveness of music. In this thesis we specifically focus on sound level, pitch contour, interval stability and pitch register. This particular subset of structural factors was selected on the basis of their relevance for communicating emotions, and their applicability in the architecture of Robin, the algorithmic composer (refer to Section 3.3). In the remaining text of the section, the emotional response related to these musical factors is discussed and is summarised in Table 2.2.

Tempo. Tempo has a major impact on arousal and a minor impact on valence. Specifically, fast tempo results in high arousal and positive valence, and slow tempo results in low arousal and negative valence (Gabrielsson & Lindström, 2010).

Mode. Mode has an influence on valence only. In particular, major mode communicates positive valence and minor mode communicates negative valence (Gabrielsson & Lindström, 2010).

Sound level. Sound level is a continuous variable that determines the volume (velocity) of the musical outcome. Sound level intensity is directly proportional to the arousal communicated in the listener. In addition, high variations of sound level may suggest negative emotions, whereas low variations tend to communicate positive emotions (Gabrielsson & Lindström, 2010).

Pitch contour (melody direction). A general agreement on the relevance of pitch contour for emotional expression does not exist. However, several studies have suggested that ascending melodies tend to be associated, among other emotions, with happiness, serenity and potency, while descending melodies are associated with sadness, vigor and boredom (Gabrielsson & Lindström, 2010).

Interval stability. Fritz and colleagues (2009) suggested that consonance is universally perceived as being more positive than dissonance. Listeners' culture and musical training do not appear to influence this.

Pitch register. High pitch register is associated with positive emotions (and at times, fear and anger), while low pitch is mostly associated with sadness (Gabrielsson & Lindström, 2010).

In addition to these structural factors, the concept of expectations was also considered, as it influences the perceived valence of music to a reasonable extent. Meyer (2008) explained that the fulfilment and the frustration of expectations impact the emotional response of the listener. According to this perspective, resolution and repetitions may suggest positive emotions, while lack of resolution can indicate negative emotions.

Table 2.2. Mapping between musical structures and the emotional dimensions of valence and arousal

		VALENCE	AROUSAL
MODE	Major	Positive	
	Minor	Negative	
TEMPO	Fast	Positive (less influential)	High
	Slow	Negative (less influential)	Low
SOUND LEVEL	High	Negative (variation)	High (intensity)
	Low	Positive (variation)	Low (intensity)
PITCH CONTOUR	Ascending	Positive	
	Descending	Negative	
INTERVAL STABILITY	Consonance	Positive	
	Dissonance	Negative	
PITCH REGISTER	High	Positive	
	Low	Negative	
EXPECTATIONS	Fulfilment	Positive	
	Frustration	Negative	

2.3.6 THE EFFECT OF EXPERTISE IN EVALUATING EMOTIONS

The extent to which emotional attribution in music is mediated by listeners' musical training has a very limited yet well-established research literature (Hargreaves & North, 2010). Since Kate Hevner (1935) found that major and minor modes were respectively mapped to happiness and sadness independently of musical training, among the studies that have addressed this matter, most of them have endorsed the opinion that the emotional perception of music is universal and not influenced by listeners' musical expertise. These studies provide support to the hypothesis that musical competence is rooted in innate predisposition, shared among the general population (Bigand & Poulin-Charronnat, 2006).

For instance, Bigand and colleagues (2005) tested in an experimental study the influence of musical expertise on emotional response to music. Participants, divided into two groups of experts (graduate music students) and non-experts (students without musical training), were asked to group 27 musical excerpts by similarity of elicited emotions. Results disclosed that expertise did not influence the subjects' emotional responses. Similar results emerged from the study of Robazza and colleagues (1994). Eighty subjects (40 children and 40 adults) were asked to rate the emotions elicited by different pieces of music. Children were divided according to their exposure to music – 20 children took music classes at school and 20 did not. Adults were divided according to their musical expertise - 20 adults had diplomas from a school of music and 20 lacked any formal musical experience. Experts and non-experts yielded similar results when evaluating the emotional connotations of music.

In contrast, other empirical research has highlighted differences due to musical training, thus supporting the implicit learning hypothesis associating musical competence to intensive musical training (Bigand & Poulin-Charronnat, 2006). Webster and Weir (2005) endorsed the idea that some perceptual differences between groups of expert and non-expert musicians exist. The objective of the study was to test whether mode, tempo and texture (i.e. music harmonisation) have an influence on the judgement of happy vs. sad music. A total of 177 college students were asked to self-report their expertise on a scale of 1 to 5. The authors treated expertise as a covariate in an ANOVA analysis, which returned significant differences in the evaluation of texture only. Those with low expertise tended to assign sadder ratings to harmonised music. Despite offering the intuition that expertise can indeed mediate music perception, this study had two limitations: (i) it relied on a categorical model, considering the emotions of happy and sad; (ii) the effect of expertise was controlled as a continuous variable, rather than selecting participants at the extreme of the distribution.

The results of a recent study conducted by Castro and Lima (2014) contrasted with previous findings. Musically trained and untrained participants divided into two age groups (N=80) were asked to rate the perceived emotions of a set of music excerpts previously validated to express four basic emotions (i.e. happiness, peacefulness, sadness and fear/threat). The results showed a correlation between length of musicians' training and sensitivity to the intended emotions. More research is required to understand the effect of musical training on humans' emotional perception of music.

2.3.7 THESIS CONTRIBUTION

This thesis aims to address some of the gaps highlighted in this review. Specifically, it provides three contributions:

1. Section 3.2 reports an empirical study on the influence of expertise in evaluating valence and arousal. The method attempts to systematically test the effect of expertise on the perception of structural factors in the circumflex model by systematically manipulating tempo and mode, and contrasting the perceptions of two groups of trained and untrained musicians.

2. Our results provide partial support to the inner predisposition hypothesis, showing that musical training does not affect the perception of arousal and valence in convergent conditions of mode and tempo manipulation. However, when the conditions diverge, trained listeners seem to be more sophisticated in understanding valence than untrained listeners, thus supporting the implicit learning hypothesis.

3. In Section 3.4, we discuss the possibility of exploiting Robin, an affective algorithmic composer, as a stimuli generator for experimental studies of music perception. An advantage offered by Robin is the possibility of generating ecologically valid music with systematic control of the tested factors.

2.4 ALGORITHMIC MUSIC COMPOSITION

One of the earliest attempts to exploit randomness for musical composition dates back to the end of the 18th century. In 1787, Mozart wrote the compositional rules of Musicalisches Würfelspiel ('Musical Dice Game') that used dice throws to compose a minuet. In essence, short sections of music were assembled according to the rolls of dice to form a minuet with 1.3×10^{29} possible combinations. Given these rules, the musicality of the resulting music relied on the coherence of the pre-composed music sections.

During the 20th century, groundbreaking scientific theories - proposed, among others, by Albert Einstein (1905) and Erwin Schrödinger (1926) - contributed to defining the influence of chaos in the physical and mathematical laws that govern the universe. Beyond their influence on science and philosophy, the concepts of chaos and randomness fascinated artists of many fields. John Cage, Iannis Xenakis and Lejaren Hiller, three of the best contemporary musicians of the last century, engaged with a number of compositions that explored stochastic processes (i.e. evolutions of random values over time) for composing music (Schwartz & Godfrey, 1993).

In the final decades of the last century, the interest in exploiting randomness in compositions reemerged due to the improved power of computational systems. Initially, mathematical models were used to manipulate timbers, often by means of additively synthesised sounds (Jacob, 1996). Gradually, computers have been used as randomising agents capable of creating unpredictable compositions through algorithmic techniques (Jacob, 1996), exploiting the mathematics of fractals, neural networks, and chaotic iterative processes (Miranda, 2001).

The improved computational power also contributed to developing algorithms capable of composing music with complex structures that are correct from a phraseological perspective (Cope, 2005). In the meantime, recent studies in music perception (Juslin & Sloboda, 2010) were combined with research on algorithmic composition techniques (Cope, 2005). This encounter produced algorithmic affective composers (Livingstone et al., 2007; Hoeberechts & Shantz, 2009). The next three subsections review the most common approaches to algorithmic composition: rule-based, learning-based, and the evolutionary approach (Todd & Werner, 1999). For a more complete review, refer to (Roads & Strawn, 1985;

Miranda, 2001). Finally, the last subsection discusses the existing algorithmic affective composers.

2.4.1 RULE-BASED APPROACH

The rule-based approach proposes to define a set of compositional rules, manually or statistically defined, that provide information to the system on how music should be created (Henz, Lauer and Zimmermann, 1996; Boenn, Brain and De Vos, 2008). Original music is generated by a number of stochastic processes driven by these rules, which can be very basic, as in the previously mentioned musical dice games by Mozart, but they can also embody complex harmonisation rules (Todd & Werner, 1999). Several rule-based algorithms are based on generative grammar, an approach to music syntax often structured with precise hierarchies of rules (Lerdahl & Jackendoff, 1985). Steedman (1984) proposed one of the most important studies exploiting the generative grammar approach. He developed a generative grammar that implements chord progressions in jazz compositions as stochastic processes.

The quality of the music generated with this approach depends substantially on the quality of human intervention. In fact, the rules have to be manually coded, and the diversity and quality of musical outcomes depends on the number of taught rules (Steedman, 1984; Wallis, Ingalls, Campana and Goodman, 2011). As a consequence, algorithm designers need to have a deep knowledge of music theory and a clear sense of their compositional goals.

2.4.2 LEARNING-BASED APPROACH

The learning-based approach proposes to reduce the reliance on human skills, instead training the system with existing musical tracks. The system is trained with existing musical excerpts and rules are automatically added (Brooks, Hopkins, Neumann and Wrigh, 1957; Hiller & Isaacson, 1957). Following this approach, Simon, Morris and Basu (2008) developed MySong, a system that automatically selects chord accompaniments given a vocal track. This study was followed by a commercial application - Songsmith – developed by Microsoft Research[7], which includes the possibility of automatically composing an entire song starting from the vocal melodies sung by the user. The system first roughly predicts the

7 Microsoft Corporation. Microsoft Research Songsmith, 2009.

notes in the vocal melody and it subsequently selects the sequence of chords that best fits the singing. A music database of 300 musical excerpts trains a Hidden Markov Model (HMM) that instructs the system with basic statistics related to chord progressions.

Another system exploiting the learning-based approach is The Continuator (Pachet, 2003). This system is ideated to provide realistic interaction with human players. The algorithm exploits Markov models to react in real time to musical input, and can learn and generate music in any style. The strengths of this system are the potential to simulate arbitrary musical styles, and the ability to produce a variety of outputs for a given input. Additionally, The Continuator makes it possible to share musical styles with other musicians, thus opening new possibilities for music collaboration while making music.

While this approach reduces the human involvement in the algorithmic composition process, the quality of music is heavily dependent on the training set. Also, this approach is not suitable when, as in our case, there is a need to have direct control of individual musical factors.

2.4.3 Evolutionary approach

Evolutionary (or genetic) algorithms are stochastic optimisation techniques loosely based upon the process of evolution by natural selection proposed by Charles Darwin (1859). In the musical domain, evolutionary algorithms have been used to create original compositions (Mitchell, 1996; Burraston & Edmonds, 2005; Gartland-Jones & Copley, 2006; Miranda & Al Biles, 2007). In most of the cases, evolutionary compositions attempt to evolve music pieces in the style of a particular composer or genre (Miranda & Al Biles, 2007). In this approach, a population of short, monophonic motifs evolves during the composition. Some systems also evolve pitches and rhythms, either concurrently or separately. Others ignore rhythm, allowing only pitch sequences to evolve, while in a few other cases only rhythm sequences evolve (Miranda & Al Biles, 2007).

In general, the evolutionary approach is particularly effective in producing unpredictable, and at times chaotic, outputs. However, the music might sound unnatural and structure-less if compared with rule-based systems, which are generally superior by virtue of the context-sensitive nature of tonal music (Nierhaus, 2009).

The evolutionary approach lacks structure in its reasoning and cannot simulate human composers' ability to develop subtle solutions to solve compositional problems such as harmonisation (Wiggins, Papadopoulos, Phon-Amnuaisuk and Tuson, 1998).

2.4.4 Algorithmic affective compositions

Recently, some studies have attempted to combine algorithmic approaches to composition with theory on music and emotion in order to automatically compose affective music. These studies have experimented with both categorical and dimensional approaches, often exploiting a rule-based approach, as it allows deeper control of individual factors.

One of the most interesting examples of an algorithmic affective composer is AMEE, a patented rule-based algorithm focused on generating soundtracks for video games (Hoeberechts & Shantz, 2009). The algorithm generates monophonic piano melodies that can be influenced in real time by adjusting the values of ten emotions with a web applet (Figure 2.5). In similar manner, Legaspi, Hashimoto, Moriyama, Kurihara and Numao (2007) adopted the categorical approach to automatically compose affective music. Using a web-based tool, users are asked to rate a set of musical scores according to a list of emotional adjectives. These ratings then

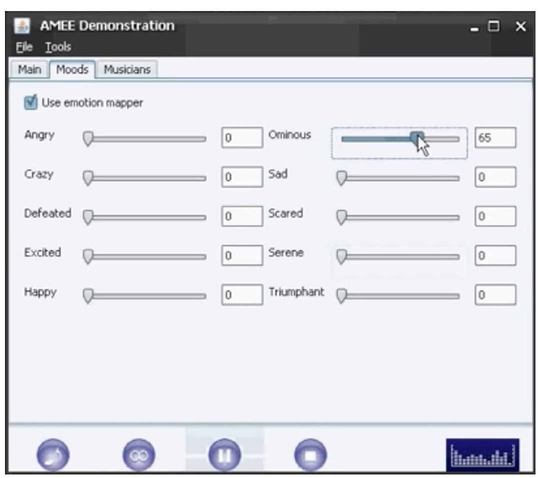


Figure 2.5 A screenshot of AMEE. The interface allows users to adjust the emotional character of the music using discrete categories of emotions.

inform the generation of rules to identify musical structures that express various affective states. In this case, the system adopts an evolutionary approach to adaptive music composition. Despite proposing interesting methods to algorithmic compositions, both systems operate using a categorical approach to emotion classification that, as discussed in Section 2.3.1, fails to address the complexity of the human emotional space.

A dimensional approach to emotion classification was adopted in three algorithmic affective composers (Livingstone et al, 2010; Oliveira & Cardoso, 2010; Wallis et al., 2011). Livingstone and colleagues (2010) followed a rule-based approach that manually collated a set of rules following related studies in music theory. The system maps emotions, which are described along the dimensions of valence and arousal, to structural and performative features (refer to Section 2.3.1). A visual interface is then provided to allow users to select the desired values of valence and arousal for the purpose of producing matching music. Likewise, Wallis and colleagues (2011) adopted an adaptive music composition system that generates piano music. A number of musical factors (i.e. mode, tempo, articulation, pitch register, sound level, voicing size, roughness, extension, and voice spacing and leading) are manipulated to match the intended emotion, which is communicated via a clickable interface that specifies the desired level of valence and arousal. Similarly, Oliveira and Cardoso (2010) proposed a system for the automatic control of emotions in music in which users can interact with the composition, providing information about the desired levels of valence and arousal. This system is based on a complex architecture that ultimately manipulates pre-composed musical scores.

These five systems have contributed to defining a novel research topic concerning algorithmic composition, allowing users to alter the emotional configuration of the composition in real time. However, a number of significant limitations reduce the practical applicability of these systems:

1. By our estimation, the quality of the music generated by these systems seems to be acceptable only when this music is considered in the context of testing the possibilities of a computer to compose affective music, rather than being enjoyable by listeners on the basis of its own merits. A formal user study that can disprove this assertion is missing from all of the reviewed literature.

2. The actual capability of the algorithms to communicate

correct emotions in the listeners has not been tested. Again, any such evaluation is absent from all reviewed studies.

3. In most of the cases, the actual interface consists of a simple applet with which users can select the intensity of discrete emotions, or values of valence and arousal. This limited utilisation, combined with the low quality of these compositions, suggests that these systems are primarily intended as pioneering explorations of a new research field, rather than serving as fully functional systems. To date, indeed, only Oliveira and Cardoso (2010) have attempted to apply their algorithm to a simple interactive installation, conducting only informal evaluations. Furthermore, this approach has the limitation of transforming pre-composed musical pieces rather than creating new music de novo.

2.4.5 THESIS CONTRIBUTION

Robin was specifically developed to overcome the limitations documented in the previous section.

1. The main objective of Robin is to generate music of a high standard, in which the control of the composition is shared with the user. To this end, a rule-based approach is utilised, as it increases the control on individual factors, to the detriment of complex musical outcomes (Section 3.3).

2. The capability of Robin to communicate correct emotions in listeners is validated with an experimental study with 33 participants (Section 3.4).

3. The potential of Robin to be used in interactive installations is tested with the case studies of The Music Room and the Twitter-Radio (Chapters 4, 5 and 6). For both installations, we also measured how much visitors liked the music, obtaining results up to standard.

Robin: An Algorithmic Affective Composer

This chapter presents Robin, an algorithmic composer that generates in real time Western classical-like music with affective connotation. To generate affective compositions, Robin follows a series of rules concerning music and emotions that were partially identified in an experimental study that is also presented in this chapter. To conclude, the chapter presents a study aimed at validating the capability of Robin to communicate expected emotional outcomes in the listeners.

3.1 INTRODUCTION

This chapter outlines the theoretical background and the implementation of Robin, the algorithmic composer that generates Western classical-like music. The objective of Robin is to allow users to interact in real time with a musical composition by means of control strategies based on emotions. Robin takes as inputs emotional information described in terms of valence and arousal (Russell, 1980), which is in turn transformed into matching music. In order for this transformation to occur, changes in the perception of valence and arousal need to be associated with alterations to structural factors in music. As reviewed in Section 2.3.4, the psychology of music suggests that the most relevant structural factors for conveying emotions are tempo and mode. Tempo has a main effect on arousal, which is proportional to tempo, and a secondary effect on valence: fast tempo tends to communicate positive emotions, while slow tempo tends to produce negative emotions. Mode has an influence on the perceived valence: major mode tends to communicate positive valence, while minor mode tends to be associated with negative valence.

An open question is whether and how this judgment varies with the listeners' expertise (Section 2.3.6). This is an important question for this thesis, given its target population of musically untrained participants, whose emotional responses to music might differ from those of an expert population. This question was tested in an experimental study (N=40) that is reported in Section 3.2. The results of the experiment, combined with findings from related work, delineate how expertise influences the perception of valence and arousal to alterations of structural factors in music. This mapping was used to encode in Robin information describing the association between variations to musical factors and changes in participants' perceived emotions. Section 3.3 presents details of this association and other characteristics of the algorithm. In particular, the operationalisation of a number of compositional rules of tonal music, and the stochastic processes that generate melody, harmony and rhythm, are detailed.

Finally, the validity of Robin is tested with an experiment that sought to determine whether the music generated by Robin communicated expected emotional responses in listeners (Section 3.4). A total of 33 participants were asked to rate the perceived valence and arousal of 20 musical snippets generated by Robin in the four conditions of 2 (positive vs. negative) valence conditions x 2 (high vs. low) arousal conditions. Results confirm that music generated by Robin adequately succeeds in communicating correct emotional responses in the listeners.

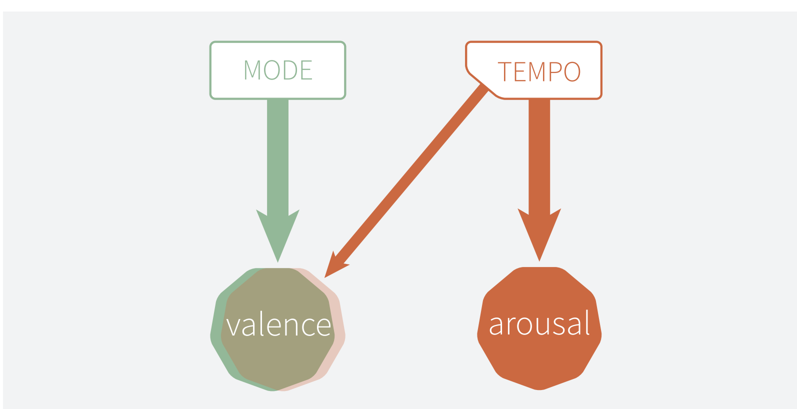


Figure 3.1. The perception of valence is influenced by the combination of mode and tempo, while the perception of arousal depends on tempo only.

3.2 STUDY I: THE EFFECT OF EXPERTISE

As reviewed in Chapter 2, the psychology of music has investigated the influence of musical factors on emotion elicitation (Section 2.3). These musical factors belong to two different categories: structural factors (e.g. tempo, mode, harmonic progression) and performative behaviours (e.g. dynamics, articulation, vibrato). The present study focuses on structural factors.

Most research credits to tempo and mode the highest relevance in terms of emotion elicitation (Gagnon & Perez 2003; Webster & Weir, 2005; Gabrielsson & Lindström, 2010). Tempo is responsible for the determining the arousal, and has a minor influence on valence. Mode is only responsible for determining the valence (Figure 3.1). The emotional response does not seem to be influenced by the listener's musical training (Robazza et al., 1994; Bigand et al., 2005), despite findings of a recent study claiming that musicians better recognise the emotional character of compositions (Castro & Lima, 2014).

A crucial issue, which was also acknowledged by (Gagnon & Perez 2003; Webster & Weir, 2005), concerns the emotional perception when mode and tempo diverge (major slow and minor fast). As reviewed in Section 2.3.4, these two studies concluded contrasting results:

- Gagnon & Perez (2003) tested the emotional perception of mode and tempo with a population of non-musicians. Results suggested that, when judging the emotional valence in diverging conditions, participants rely more on tempo than mode; thus, the minor fast condition is perceived as more positive than the major slow condition.
- Webster & Weir (2005) tested the emotional perception of mode, tempo and texture with a mixed population of musicians and non-musicians. Results suggested that musical pieces in diverging conditions have similar, neutral values.

We argue that the different results might be attributable to the different populations involved (non-musicians only vs. both levels of expertise). Specifically, the participants of the former study might have attributed less importance to mode because non-musicians show poor discrimination between modes (Halpern, Martin, and Reed, 2008) Furthermore, both studies measured listeners' perceptions simply by judging happy vs. sad conditions (Gagnon & Perez 2003; Webster & Weir, 2005).

The present study proposes to further examine this issue, measuring participants' emotional responses in the two dimensions of valence and arousal. The author's hypothesis is that the perception of mode and tempo might be influenced by listeners' musical knowledge. Musicians, who clearly perceive the difference between modes, may use this information to a larger extent than non-musicians to rate the emotional meaning of music. In order to test this hypothesis, an experiment was conducted with 40 participants equally distributed into two groups of musically trained and untrained participants.

3.2.1 EXPERIMENTAL HYPOTHESES

H1. Tempo is a musical property that is easily recognisable by all listeners, as opposed to differences of mode, which are better understood by trained listeners (Halpern et al., 2008). Consequently, we expect trained musicians to employ this knowledge to a greater extent when judging perceived valence, thus rating the valence of music in major mode with higher scores than that of music in minor mode, as opposed to musically untrained listeners, who are expected to differentiate between modes to a lesser extent.

H2. In the case of diverging conditions of mode and tempo (i.e. major-slow and minor-fast), given that mode and tempo operate in opposite directions, listeners might have conflicting feelings. Following H1, expertise may have an influence on this result. Musicians are expected to be mainly influenced by mode; thus, they are expected to rate valence higher in major-slow than in minor-fast. By contrast, non-musicians, who are less familiar with mode, might mostly employ tempo information thus rating minor-fast as positive and major-slow as negative (Table 3.1).

Table 3.1 Hypothesised interaction between mode and tempo in the perception of valence.

		Fast tempo	Slow tempo
NON-MUSICIANS	Major mode	+	-
	Minor mode	+	-
MUSICIANS	Major mode	+	+
	Minor mode	-	--

3.2.2 DESIGN

The hypothesis of this study was tested in a 2x2 within-subject experiment. Mode was manipulated in major and minor conditions, and tempo was set to 80 BPM and 160 BPM. The dependent variables of the experiment were perceived valence and arousal.

3.2.3 STIMULI

A professional composer *ad-hoc* composed seven short musical excerpts. The emotional connotation of the music was kept as neutral as possible following common compositional strategies, such as identical harmonic progression. The snippets consisted of an accompaniment and a solo line, each organised in four bars of pseudo-ecological music (i.e. each could not be considered a song by itself but could potentially be the first bars of a musical piece). To impart snippets with an ecological validity, the solo line was composed to be meaningful. Each snippet was manipulated in a minor and major mode version. Snippets were also manipulated with respect to tempo, generating two versions for each mode-manipulated snippet. This manipulation yielded 7 variations of 2 (mode: minor vs. major) × 2 (tempo: slow vs. fast) snippets for a total of 28 snippets.

The scores of 28 snippets can be found in the Appendix, while the mp3 files are available at http://bit.ly/1EypafP

The fast version played at 160 BPM with a high density of notes in the accompaniment, while the slow version played at 80 BPM with a low density of notes. Note density was also varied as previous studies suggested that music with a high density of notes is generally perceived as faster as compared with a piece having identical BPM but lower density of notes (Gabrielsson & Lindström, 2010). To operationalise this distinction, the difference between the two different tempo conditions was increased using eighth notes in the accompaniment in the 160 BPM pieces and quarter notes in the 80 BPM pieces. For simplicity, from this point forward we will refer to *tempo* as comprising note density, as well.

All of the snippets had the same harmonic progression (i.e. I - IV - V - I, one of the most common progressions in tonal music such as Baroque, classical, romantic and pop). The snippets were composed in different key signatures to increase the generalisability of the stimuli. All of the melodies had the same pitch range, as relevant deviations in the range of the pitch might impart an emotional influence (Gabrielsson & Lindström, 2010). The intervals that were used varied within an octave, as ranging over the octave may also impact the emotional reaction (Gabrielsson & Lindström, 2010).

3.2.4 Procedure

Participants were recruited from students and staff of the University of Trento and the Conservatory of Trento. A total of 40 participants took part in the experiment. Twenty participants (10 F) were trained musicians with at least five years of music school (or comparable institutions); the remaining 20 (5 F) had no formal music education. The age of participants ranged from 19 to 42 years with an average age of 24.7 years. Most participants (N=30) were Italian, while the rest were originally from different European and Asian countries. Each experimental session ran in a silent room at the Department of Information Engineering and Computer Science at the University of Trento, Italy. Participants sat in front of a laptop listening to the auditory stimuli through a pair of AKG K550 headphones (Figure 3.2).

Figure 3.2. The experimental setup. A participant is asked to rate the perceived valence on a scale from 1 (negative) to 7 (positive). He communicated the intended value operating on an external number pad keyboard.

Before starting the experiment, each participant received detailed instructions by means of written notes (in English and Italian). Participants were initially presented with four training excerpts in order to become familiar with the interface and the task. The 28 snippets were presented in a random order, which differed among participants. While snippets were played, the display was completely white. At the end of each snippet, participants were prompted with a screen asking to report what emotion that particular music had communicated. In particular, they were asked to independently rate valence and arousal on bipolar scales from 1 (*negative* or *relaxing*) to 7 (*positive* or *exciting*). To assign the desired value of valence and arousal, they interacted with the system through a USB mini number pad keyboard. Between each listening, the computer played a sequence of random notes as the screen turned to black. These random-note sequences have been previously validated for masking the effects of previously played music (Bharucha & Stoeckig, 1987). Figure 3.3 presents a timeline of the experimental session.

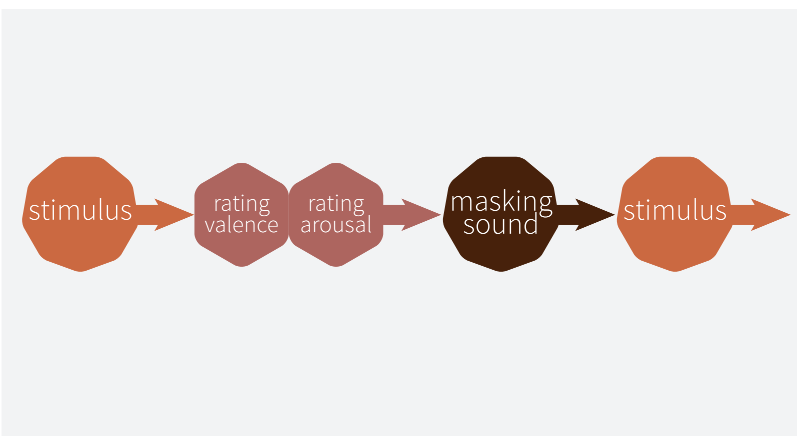


Figure 3.3. Timeline of the experimental session: 1) the participant first listens to a stimulus; 2) she/he is prompted to rate the perceived valence and arousal of a piece; 3) a masking sound is played; 4) another stimulus is presented to the participants.

At the end of the experiment, in order to assess the homogeneity of the original snippets, participants were asked to indicate how much they like each of the seven original snippets from 1 to 7, with 1 representing "*not at all*" and 7 "*a lot*". The snippets, which were presented in major mode and at 120 BPM, were presented in random order by the computer.

3.2.5 RESULTS

Data gathered from the question investigating the liking of the seven original stimuli showed that all of participants reported similar scores. The average values for each snippet varied from 3.45 to 3.83, thus proving that participants did not have particular implicit preferences for the adopted stimuli. Thus, valence and arousal ratings were computed by averaging scores to the seven stimuli.

Subsequently, a repeated measures ANOVA was performed on valence and arousal ratings separately. In both cases, *mode* (major and minor) and *tempo* (fast and slow) were the within-subject factors, and *expertise* (low and high) was the between-subject factor. A repeated measures ANOVA was also performed on the data describing the level of liking of the seven snippets (within-subjects) compared with the two categories of expertise (between-subjects). Here, we use a *p* level of .05 for all statistics, and we report all analyses that achieve these levels of significance.

Valence. The analysis showed a significant main effect of *mode* [$F(1,38) = 279$, $p<.001$] and *tempo* [$F(1,38) = 106.6$, $p<.001$] on valence. Major mode was associated with high valence (mean 5.05, SD .78)

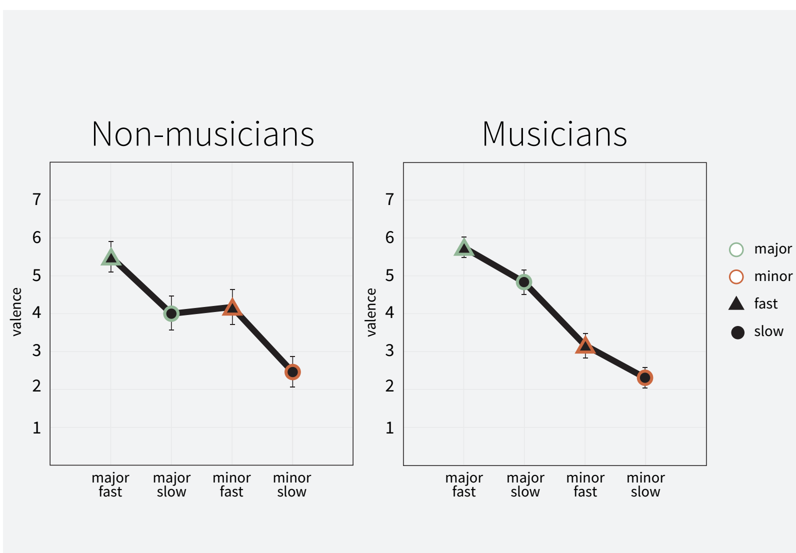


Figure 3.4 Average values and ST of valence divided by expertise. The perceived valence varies with different levels of expertise in case of diverging conditions of mode and tempo.

and minor mode with low valence (mean 2.97, SD .87). In addition, fast tempo was associated with high valence (mean 4.65, SD 1.29) and slow tempo with low valence (mean 3.37, SD 1.27). Figure 3.4 illustrates these results. The interactions between *mode* and *expertise* ($F_{(1,38)}$=27.6, p<.001) and *tempo* and *expertise* ($F_{(1,38)}$=10.9, p<.001) were also significant. Trained listeners assigned higher scores to major-mode snippets than to minor-mode snippets, regardless of whether they were convergent or divergent. The emotional response of non-musicians was less sophisticated. Snippets in divergent conditions were both perceived as having neutral valence. The major-fast condition was rated with high scores while the minor-slow condition received low scores.
Arousal. No significant difference between major and minor mode, nor any impact of expertise emerged. The analysis of arousal showed a significant main effect for mean scores of *tempo* ($F_{(1,38)}$= 311.3, p<.001). Fast tempo was associated with high arousal (5.02, .80) and slow tempo with low arousal (2.50, .83). Figure 3.5 presents the arousal for the four conditions according to expertise. No significant interactions were found.

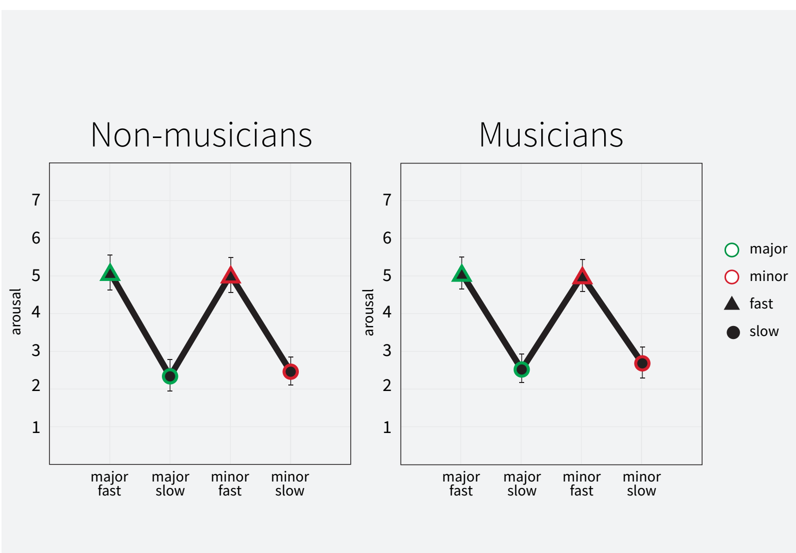
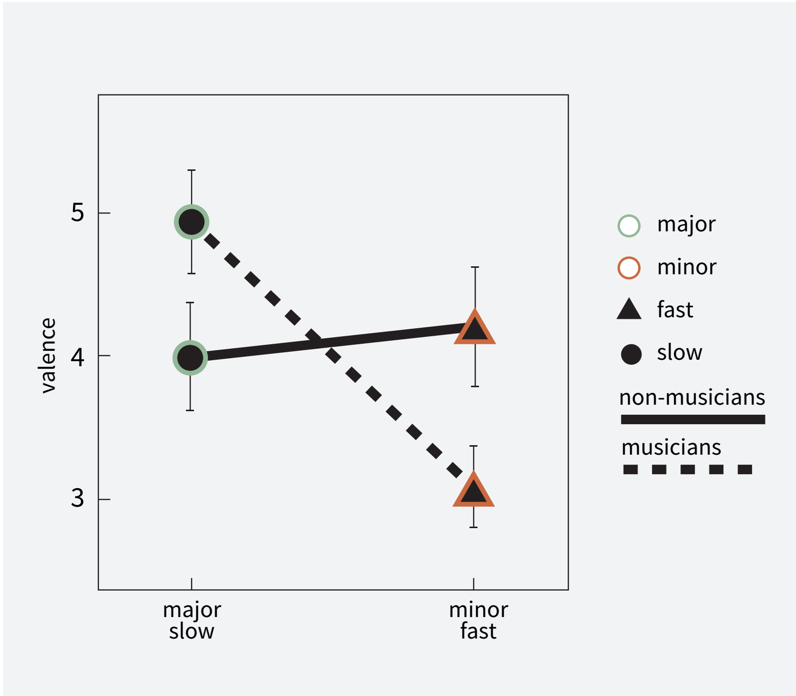


Figure 3.5. Average values and ST of arousal divided by expertise. The perceived arousal does not vary with different levels of expertise.

Figure 3.6. The effect of expertise in the evaluation of valence in divergent conditions. Trained musicians employed mode more than tempo to evaluate the valence, while non-musicians rated the two conditions with similar, neutral values.



3.2.6 DISCUSSION

The results of this experiment suggest that expertise has an impact on listeners' emotional response to music. In particular, expertise mediates the impact of mode on ratings of valence, but not that of arousal. Non-musicians appeared to have been influenced by the combined impact of mode and tempo. When the information was divergent, they rated valence halfway between the two convergent extremes, thus confirming the findings from Webster and Weir (2005) and contrasting the findings of Gagnon and Perez (2003). On the other hand, in these conditions, musicians relied on mode above tempo when judging a piece of music on its valence, disputing results from Webster and Weir (2005), who indicated that expertise does not have a significant effect on valence perception. Figure 3.6 illustrates that in divergent conditions, musicians assigned more importance to *mode*, while non-musicians did not take a clear position. The results of listeners' perceptions of valence are shown in Table 3.2. Arousal, by contrast, seemed to be related only to *tempo*, disregarding *mode* and *expertise*. This result confirms past findings from (Gomez & Danuser, 2007; Juslin, 1997).

Table 3.2 Tested interaction between mode and tempo in the perception of valence.

		Fast tempo	Slow tempo
NON-MUSICIANS	Major mode	+	~
	Minor mode	~	-
MUSICIANS	Major mode	++	+
	Minor mode	-	--

In summary, the acquisition of musical expertise appeared to have an influence on the emotional experience of people listening to music. As people participate in formal musical training, they become particularly sensitive to mode when evaluating the emotional character of music. In particular, in the case of divergent conditions of mode and tempo, trained musicians attributed mode a higher relevance when evaluating the valence of a musical piece. By contrast, non-musicians seemed to be unable to identify a clear emotional character in this condition.

This result is particularly important for the research objective of this thesis, as it clarifies the emotional response of non-musicians to changes of mode and tempo, the structural factors most significantly responsible for communicating emotions. This finding, combined with the results of related works, offers a solid bases upon which to build an algorithmic affective composer.

3.3 DEVELOPMENT

This section introduces Robin, an affective-based algorithmic composer designed to make the experience of musical creativity accessible to all users. Rather than providing information in a musical language, users interact with the composition by means of control strategies based on emotions. Specifically, they communicate the intended values of valence and arousal. To ensure consistency with user interaction, the system continuously monitors input changes, adapting the music accordingly via the manipulation of seven musical factors (Section 2.3.5).

Robin is specifically designed to be used in interactive installations targeting a general population. To this end, two requirements must be met: i) the composition has to adapt in real time to user input; ii) the generated music must be understandable even by an untrained audience. The first requirement led to the adoption of a rule-based approach to algorithmic composition. This approach is particularly suitable for this study, as it guarantees accurate control of the compositional process given that rules are manually coded. As explained in Section 2.4.1, this approach has the drawback of requiring human intervention; thus, the outcome depends on the quality of the taught rules. To this end, a professional composer was continuously involved during the design and testing of the algorithm to guarantee high standards of the compositions. The second requirement led to the adoption of tonal music, given its potential to reach a wider audience. Tonal compositions are indeed ubiquitously present in Western culture, so that even those who lack musical training internalise the grammar of tonality as a result of being exposed to it (Winner, 1982).

3.3.1 SYSTEM ARCHITECTURE

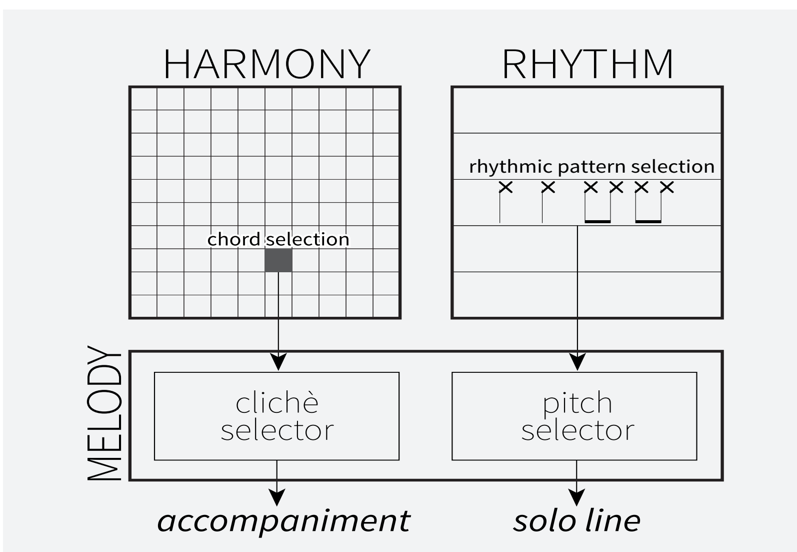


Figure 3.7. The architecture of Robin, the algorithmic affective composer.

Robin generates scores composed of a solo and an accompaniment line. The process of score generation is grounded upon a number of compositional rules of tonal music driving stochastic processes, which in turn generate *harmony, rhythm* and *melody* (Figure 3.7). The *Harmony* module determines the chord progression by following a probabilistic approach. The selected chords are then fed into the *Melody* module. Here, it combines with (i) a *rhythmic* pattern that is completed with pitches from the scale and outputs the solo line; (ii) a cliché selector that outputs the accompaniment line.

3.3.2 HARMONY

Traditionally, harmony is examined on the basis of chord progressions and cadences. Following previous works (Steedman, 1984; Nierhaus, 2009), the transition probabilities between successive chords are defined as Markov processes. Chord transition data can be collected by analysing existing music, surveying music theory, or following personal aesthetic tastes and experiences (Chai & Vercoe, 2001).

In Robin, correlation of chords does not depend on previous states of the system. A first-order Markov process determines the harmonic progression as a continuous stream of chords. The algorithm starts from a random key and then iteratively processes a Markov matrix to compute the successive chords (Table 3.3). The 10 x 10 matrix contains the transition probabilities among the degrees of the scale. The entries are the seven degrees of the scale as triads in root position, and three degrees (II, IV, V) set in the VII chord. The transition probabilities are based on Piston's (1941) study of harmony. At each new bar the system analyses the transition matrix and selects the degree of the successive bar: the higher the transition value, the higher the probability to be selected. For instance, being VII the current degree of the scale, the I degree will be selected as successive chord in the 80% of the cases, on average, whereas the II7 will be selected in the 20% of the cases.

In addition, in order to divide the composition into phrases, the system forces the harmonic progression to a cadence (a conclusion of a phrase or a period) every eight bars. Finally, as to generate compositions with more variability, Robin can switch between different keys and perform V and IV modulations.

Table 3.3. Transition probability matrix among the degrees of the scale.

	I	II	III	IV	V	VI	VII	IV7	V7	II7
I	0	0.05	0.05	0.30	0.20	0.05	0.1	0.05	0.15	0.05
II	0.04	0	0.04	0.04	0.45	0.08	0	0	0.35	0
III	0	0.07	0	0.21	0.07	0.65	0	0	0	0
IV	0.15	0.10	0.05	0	0.35	0.05	0	0	0.30	0
V	0.64	0.05	0.05	0.13	0	0.13	0	0	0	0
VI	0	0.40	0.10	0.10	0	0	0	0	0	0.40
VII	0.8	0	0	0	0	0	0	0	0	0.2
IV7	0	0.30	0	0	0.30	0.30	0	0	0.10	0
V7	0.9	0	0	0.05	0	0.05	0	0	0	0
II7	0	0	0	0	0.5	0	0	0	0.5	0

3.3.3 RHYTHM

At each new bar, the *Rhythm* module randomly selects a new rhythmic pattern from a set of several dozen of patterns. All possible combinations of rhythms composed of *whole, half, quarter, eight* and *sixteen notes* are available. In addition, the same combinations of notes in *triplets* are available. The harmonic rhythm is one bar long, and each bar has a time signature of 4/4.

3.3.4 MELODY

Once selected, the rhythmic pattern is filled with suitable pitches in the *Melody* module. This process is performed in three steps:

1. The *pitch selector* receives the rhythmic pattern and the current chord (Figure 3.8.a).

2. All of the significant notes in the bar are filled with notes of the chord. The notes regarded as significant are those whose

Figure 3.8. Melody notes selection. a) The pitch selector receives the rhythmic pattern and the chord. b) The relevant notes of the melody are filled with notes of the chord. c) The remaining spaces are filled with notes of the scale to form a descending or ascending melody.

duration is an eighth note or longer or that are at the first or the last place (Figure 3.8.b).

3. The remaining spaces are filled with notes of the scale. Starting from the leftmost note, when the algorithm encounters an empty space, it checks the note on the left and it steps one pitch up or one pitch down, depending on the value of the melody direction (Figure 3.8.c). The accompaniment line is selected at each new bar. A number of clichés (accompaniment typologies) are available. The clichés essentially differ in the density of the notes in the arpeggio. Each cliché defines the rhythm of the accompaniment, and the notes of the accompaniment are degrees of the chord.

3.3.5 Definition of high-level musical structures

As opposed to similar affective composers such as AMEE (Hoeberechts & Shantz, 2009), Robin does not allow the definition of high-level musical structures like verses and sections. Human composers often make wide use of high-level structures to create emotional peaks, or to develop changes in the character of the composition. Therefore, including such structures in a real-time algorithmic composer is not a viable solution, as user interaction with the system cannot be predicted in advance.

AMEE deals with high level musical structures by introducing forced abortion in the process of music generation (Hoeberechts & Shantz, 2009). However, this design solution causes dramatic interruptions, thus reducing both musical coherence and a natural evolution of the composition itself. To this end, the only high-level structural elements composed by Robin are short theme repetitions, which partially simulate choruses and verses, and cadences, which define phrases.

3.3.6 Operational definition of emotion in music

The structural factors manipulated by Robin to infer changes in the communicated emotions, defined in terms of valence and arousal, are: *tempo, mode*, *sound level*, *pitch contour*, *interval stability*, *pitch register* and *expectations* (refer to Section 2.3.5). This section discusses how the alteration of the emotional response of the seven musical parameters is operationalized in Robin.

Tempo. Robin treats tempo as a continuous variable measured in BPM. Besides BPM, Robin manipulates *note density* by selecting rhythmic patterns and accompaniment clichés with appropriate note density.

Mode. Robin supports the change between modes in the *Harmony* module, where the chords transition probability matrix is populated with notes based on the selected mode.

Sound level. Sound level changes by manipulating the *velocity* of the MIDI.

Pitch Contour (Melody Direction). Robin determines the direction of the melody using the method described in Section 3.3.4.

Interval stability. Dissonance is achieved by inserting a number of out-of-scale notes in both melody and harmony. For the current implementation of Robin, the distance between the out-of-scale notes and the scale is not considered.

Pitch Register. The pitch register centre of the compositions generated by Robin ranges from C2 (lowest valence) to C5 (highest valence).

Expectations. Robin operationalises expectations repeating themes and recurring patterns that the listener quickly comes to recognise as familiar. Table 3.2 schematises how the combination of these elements determines the desired emotional expression in terms of valence and arousal.

3.3.7 DISCUSSION

The development of Robin required almost two years of continuous iteration with a composer who was involved in the design of the algorithm since its early stages. A number of professional musicians were also involved in several focus groups aiming to evaluate the quality of the music and to suggest possible improvements. In general, Robin was met with enthusiasm by the musicians. The possibility of influencing the composition in real time by communicating the intended emotions was particularly appreciated.

The final version of the algorithm received an even stronger interest. Every musician agreed that the quality of the music was satisfying as the generated product of an algorithmic composer. A common feedback suggestion from several musicians was to include expressive performance behaviour to improve the humanness of the music. This is certain to be one of the future works addressed for Robin, whose real-time affective score generation capability should be easily combined with existing systems for the automatic modeling of expressive contents (Canazza, De Poli, Rodà, 2014).

Despite the visible success of Robin, a more formal evaluation was required to test the quality of the algorithm. Specifically, given the role of Robin relative to the large perspective of this thesis, its effective capability of communicating expected emotions in humans is of fundamental importance.

3.4 STUDY II: VALIDATION

In order to test the capability of Robin to communicate specific emotions in the listeners, an experimental study was conducted. The objective of the study was to investigate whether Robin was able to communicate a predictable amount of valence and arousal in the participants. Participants were asked to listen to a number of pieces generated by Robin in different emotional conditions, and to self-report the perceived level of valence and arousal. The experiment could be declared successful if the intended levels of valence and arousal were correctly identified by participants. In other words, Robin was confirmed to communicate correct emotions in the listeners if:

1. Listeners perceived positive (negative) valence when Robin was configured to generate music with positive (negative) valence;
2. Listeners perceived high (low) arousal when Robin was configured to generate music with high (low) arousal.

3.4.1 DESIGN

Robin's capability to communicate specific emotions was tested with four combinations of valence (positive vs. negative) and arousal (high vs. low) in a 2*2 within-subjects design. For each condition, Robin generated five different piano snippets (30 seconds long), for a total of 20 snippets. The snippets consisted of an accompaniment and a solo line. Once generated, a 3-second fade-out effect was added at the end of each snippet[1].

1 Can be found at http://bit.ly/1HSKjOl

3.4.2 STIMULI

In order to generate music with affective flavours in the four conditions, Robin manipulated mode, tempo, sound level, pitch contour, pitch register and expectations. All of these factors, excepting tempo, influence either valence or arousal. *Tempo*, on the other hand, has a major effect on arousal, but it also influences valence (Section 2.3.4). This secondary effect is particularly evident for untrained musicians, as revealed in the study described in Section 3.1. The double influence of tempo was operationalised as follows:

1. Snippets with high arousal were twice as fast as than snippets with low arousal;
2. Snippets with high valence were slightly faster (8/7 times) than the snippets with low valence.

Table 3.4 shows the mapping between the six factors and the four conditions of valence/arousal (**+-** = positive valence / high arousal, **+-** = positive valence / low arousal, -**+** = negative valence / high arousal, -- = negative valence / low arousal).

Table 3.4. The value of the six factors in the four different conditions of valence and arousal.

	++	**+-**	**-+**	**--**
Mode	Major	Major	Minor	Minor
Tempo (BPM)	160	80	140	70
Sound level	High	Low	High	Low
Contour	Ascending	Ascending	Descending	Descending
Octave	High	High	Low	Low
Repetitions	Yes	Yes	No	No

The double action of tempo on both dimensions might have side effects in this experimental design. Specifically, low tempo might reduce the perceived valence of the +- condition, and high tempo might increase the perceived valence of the -+ condition. The hypotheses of the study are listed in Table 3.5.

Table 3.5. The expected values of valence and arousal. In the valence graph, the side effect of tempo in the two diverging conditions might cause lessened responses.

		High arousal	Low arousal
PERCEIVED VALENCE	Positive valence	+	+
	Negative valence	-	-
PERCEIVED AROUSAL	Positive valence	+	-
	Negative valence	+	-

3.4.3 PROCEDURE

Participants were recruited from students and staff of the University of Trento, Italy. A total of 33 participants took part in the experiment. Similarly to the previous experiment reported in Section 3.2, sessions ran in a silent room at the Department of Information Engineering and Computer Science at the University of Trento, Italy. Participants sat in front of a monitor wearing AKG K550 headphones. Before starting the experiment, participants were informed about the task they had to complete by means of written notes (in English and Italian). They were initially presented with four training excerpts in order to become familiar with the interface and the task. Then, the 20 snippets were presented in a random order.

In order to measure valence and arousal separately, participants were asked to rate them on two semantic differential seven-point scales, from 1 (negative or relaxing) to 7 (positive or exciting). In addition, they were asked to indicate, from 1 to 7, how much the liked each snippet (*liking*). To assign the desired value of valence, arousal and likeness they typed the numbers 1-7 on a keyboard. Between each listening, the computer played a sequence of random notes; each of the random snippets was randomly selected from a set of five pre-recorded 15-seconds snippets composed of random notes.

3.4.4 RESULTS

A two-way within-subjects ANOVA analysis was performed on valence, arousal and liking ratings separately. In both cases, valence (high and low) and arousal (high and low) were the within-subject factors. To disambiguate between the intended valence and arousal (independent variables) and the tested valence and arousal (dependant variables), we will refer to the first couple as *intended* and the second couple as *perceived.*

Here, we use a *p* level of .05 for all statistics, and we report all analyses that reach these levels. The average values and SD of perceived valence and arousal are illustrated in Table 3.6.

Table 3.6. Means and standard deviations for the three tested factors, in the four emotional conditions.

Intended combination	Perceived valence		Perceived arousal		Likeness	
	Mean	*SD*	*Mean*	*SD*	*Mean*	*SD*
++	5.206	.5711	5.315	.7072	4.800	.8874
+-	4.261	.6528	3.527	.8409	4.182	.8809
-+	3.915	.8646	4.952	.6226	4.339	.9320
--	3.224	1.263	2.521	1.166	4.182	1.211

Perceived Valence. The analysis of the means revealed that **++** scored the highest value of valence (5.21), and -- mean scored the lowest value of perceived valence (3.22). The expected side effect of tempo was observed: **+-** and **-+** resulted in similar neutral scores (4.26 and 3.91, respectively). The analysis of the perceived valence showed a significant main effect both for intended valence [$F(1,32) = 32.90$, $p<.001$] and for intended arousal [$F(1,32) = 36.8$, $p<.001$]. The interaction between the two factors was not significant. This result indicates that the manipulation of intended valence and intended arousal contributes to defining the perception of valence, but that the two factors do not intersect. Specifically, in the same intended arousal conditions, the positive intended valence snippets reported more positive values, and in the same intended valence conditions, the high intended arousal snippets reported higher values.

Perceived Arousal. The means of the four conditions matched our expectations: the two high intended arousal conditions scored high values of perceived arousal (5.31 and 4.95 for positive valence and negative valence conditions, respectively), and the two low intended arousal conditions scored low values (3.52 and 2.52 for positive valence and negative valence conditions respectively). The analysis of the perceived arousal showed a significant main effect for both intended valence [$F(1,32) = 29.4$, $p<.001$] and for intended arousal [$F(1,32) = 147.9$, $p<.001$]. The interaction between the two factors was also significant [$F(1,32) = 12.6$, $p<.005$]. These data suggest that the manipulation of both intended valence and intended arousal contribute to defining the perception of arousal, and that their intersection also had a consequence. The effect of the interaction between the two factors is evident in the difference between the **+-** and -- conditions: in the case of low intended arousal, the perceived arousal was significantly higher when combined with positive intended valence.

Liking. The means of participants' ratings for each snippet varied between 3.72 and 5.15 with an average of 4.38. The ANOVA analysis revealed that intended arousal was the most significant factor with respect to liking [$F(1,32) = 8.978$, $p<.01$]. The interaction effect of arousal and valence was also significant [$F(1,32) = 4.735$ $p<.05$]. The favourite condition was high valence combined with high arousal (mean 4.80, SD .88), while valence had no implication on low arousal conditions, which reported identical values (4.18, 1.21).

3.4.5 Discussion

The experiment showed that listeners' emotional responses to the music composed by Robin met our expectations to a significant extent. The perceived arousal perfectly matched the intended arousal. The perception of valence, however, matched the intended valence only when the conditions converged. In the case of diverging conditions the perceived valences of -+ and +- reported similar, neutral averages (Table 3.7).

Table 3.7. Measured levels of perceived valence and arousal.

		High arousal	Low arousal
PERCEIVED VALENCE	Positive valence	+	~
	Negative valence	~	-
PERCEIVED AROUSAL	Positive valence	+	-
	Negative valence	+	-

This finding can be explained with reference to the difficulty of non-musicians to distinguish divergent emotional stimulation (Section 3.2.4). A possible solution would be to decrease tempo in the –+ condition or to increase it in the +- condition to increase the difference of perceived valence between the two conditions. Rebalancing tempo values would have indeed produced stronger resuts. For instance, a related study from Bresin and Frieberg (2011) suggested that that happy performances are usually played almost 4 times faster than sad performances.

To summarise, the result of the experiment suggests that the current implementation of Robin correctly communicate arousal level in the listeners but it fails to communicate (i) very positive valence when combined with low arousal and, (ii) very negative valence when combined with high arousal. To address this issue for future implementations of the system, besides adjusting tempo level, we will consider including a number of performative behaviours that could influence the perceived emotions. Related research on this topic proved particularly effective for communicating expected emotional response in the listeners (Juslin & Sloboda, 2010).

3.5 CONCLUSION

As I started working on this Ph.D. project, affective-based algorithmic composers had only been developed since very recently (Legaspi et al., 2007; Hoeberechts & Shantz, 2009; Livingstone et al., 2010; Wallis et al., 2011). Initially, I considered integrating one of these systems into the framework presented in this thesis. This would have allowed me to focus my work on the interaction design aspect only. However, I soon realised that the quality of the music generated with those systems was inadequate. Furthermore, the tunes that these systems generated did not match my personal aesthetic. This is an important issue that would require being taken into much deeper consideration when discussing the research taking place where art and science overlap. In fact, in addition to pure scientific methodologies and techniques, aesthetics and tastes play a crucial role in the evaluation of such a system, which can potentially be flawless from a methodological and mathematical point of view, but still incapable of satisfying the aesthetics of the listeners.

Following this line of thought, personal aesthetics were also considered when integrating Robin in The Music Room and in the TwitterRadio, which will be fully discussed in the successive chapters. For these interactive installations in fact, besides the structural factors discussed in this chapter, different instrumentations were mapped to different affective states by taking into account personal taste and intuition. In these installations, when very positive valence situations occur, violins performing in staccato articulation double the solo line. By contrast, in the case of very negative valence, either trilling violins or low and deep horns are triggered. This choice, which was grounded both on personal tastes as well as on related work findings (Juslin & Sloboda, 2010; Eerola, Friberg and Bresin, 2013), was particularly appreciated by the audience of the installations.

Another interesting aspect emerged from the studies presented in this chapter. Robin revealed its potential of being employed as a composer of snippets in experimental studies testing the perception of structural factors. This statement finds confirmation from the results of the study presented in Section 3.4. Not only did Robin prove to be successful in generating music with precise emotional connotations, but also the average listener's enjoyment of the snippets even exceeded that generated by a human composer, in the first study.

Interface Design: The Music Room

THIS CHAPTER REPORTS THE DESIGN AND THE TECHNICAL IMPLEMENTATION OF THE MUSIC ROOM, AN INTERACTIVE INSTALLATION THAT ALLOWS VISITORS TO INFLUENCE THE EMOTIONAL ASPECT OF AN ORIGINAL CLASSICAL-LIKE MUSICAL COMPOSITION BY MOVING THROUGHOUT A ROOM.

4

4.1 INTRODUCTION

The first interface to Robin in the context of interactive installation was The Music Room. In this installation, the suitability of emotions as mediators of music complexity, and of movements as actualizations of this mediation is tested. Replacing musical notations with emotions in the process of music making results in a number of critical implications in the design process.

1. In order to be communicated to the system, emotions need to be encoded into specific media. One suitable medium through which emotions can be conveyed is *bodily movement*, as it is naturally associated to music and emotions across different cultures (Sievers, Polansky, Casey and Wheatley, 2013).

2. The involvement of the player changes: the traditional paradigm based on a note-to-note control is replaced by control strategies based on the emotion the player intends to convey. This change has a major impact on the experience of the player, which should be borne in mind when designing such interfaces.

3. An algorithmic composer needs to be developed that systematically converts user input described in emotional language into compositional rules, which are in turn used to direct the composition. This issue was usefully described by Jacob (1996) when describing the issue of designing computational systems aiming at reproducing human ability to make music: “In short, creativity comes in two flavors: genius and hard work. While the former may produce more “inspired” music, we do not fully understand it and therefore have a slim chance of reproducing it. The latter resembles an iterative algorithm that attempts to achieve some optimal function of merit, and is therefore more easily realizable as a computer program.”

In The Music Room, an interactive installation in which where visitors can direct the emotional character of music by means of their movements. The installation was intended to be experienced by dyads of visitors interacting with each other by moving throughout a room. The interaction paradigm is that of intimacy: the more proximal the visitors are, the more positive the music. The intensity of the composition is also left to visitors' control: the faster they move, the louder and faster the music. An abstract of the functionality of The Music Room follows:

> *A couple of visitors enters the room and music starts playing. Once in the room, they can direct the emotional aspect of the music by interacting with each other while moving throughout the space. Their relative distance and average speed determine the emotional character of the music generated. Specifically, their relative distance influences the pleasantness of music (valence) and their average speed its intensity (arousal). The music sounds positive and romantic when they stand close; it gradually becomes sadder as they move apart; and it sounds totally unpleasant when they are on the opposite sides of the room. The intensity of the music is high when they move fast, running or dancing; it calms down as they move slower; and it almost pauses when they freeze.*

This idea evolved through two years of research through design, a practice intended to exploit design to produce knowledge (Fallman, 2003; Zimmermann et al., 2007). This design strategy fit particularly well the objectives of this thesis, which aimed at exploring new solutions of music making via design. The two years of project development involved a conceptual design stage (i.e. transforming requirements into a conceptual model: Sharp, Rogers and Preece, 2007), which emerged through the application of the PACT framework (Benyon, Turner and Turner, 2005), This framework helps designers to investigate the design process by means of a user-centred technique based on *Peoples*, *Activities*, *Contexts* and *Technologies*. *Peoples* defines the target audience that actively or passively engages with the interface. The target audience of this work was already identified:

dyads of visitors who are not necessarily trained in music. The other three entities of the PACT framework are discussed later in the chapter. In particular:

1. Section 4.2 discusses the *Activities*, the main purpose of the system, i.e. the interaction scenarios as envisioned by the designer.

2. Section 4.3 elaborates the *Technologies*, i.e. the actual implementation of the hardware and software components necessary to execute the installation.

3. Section 4.4 probes into the *Contexts* of the installation, the environmental elements of the system.

4.2 ACTIVITIES

During the conceptual design phase, three basic scenarios were envisaged, *dancing*, *composing* and *acting*. The scenarios were enriched with graphics and storyboards (Figure 4.1) and used as design probes in a workshop involving 12 user experience researchers. The participants of the workshop supported the conceptual idea of The Music Room. Also, a number of interesting considerations emerged during the workshop. In particular, participants were keen to discuss the possible behaviours of the people in the room – whether they would be more interested in creating music or enjoying the intimacy with the partner or a friend. Next, we report the three scenarios as elaborated following the workshop.

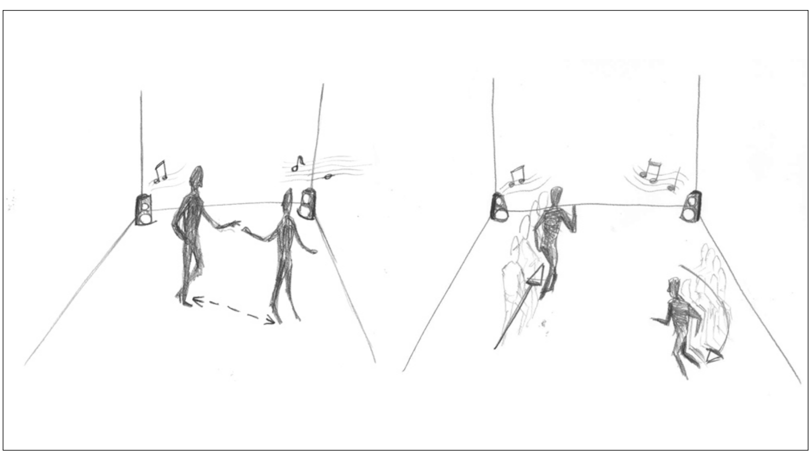


Figure 4.1. Sketched scenarios of The Music Room presented at the design workshop. In the left panel the dyads are dancing closer slowly. In the right panel they are running at a distance.

4.2.1 DANCING SCENARIO

Andrea and Adriano are visiting an art museum. Attracted by the music and by the posters advertising The Music Room, they decide to try it. As they enter the room, they start moving rhythmically to the music. Initially, they freely dance in the room individually. Then, they decide to test the system by trying different combinations of distance and speed while keeping engaged with dancing. Once they realise that they can indeed successfully influence the emotionality and the intensity of the music, they move close together, performing some ballet dancing while generating a romantic classical music. Then, they gradually increase the speed of their movements to eventually dance frantically, far from each other, while generating some experimental jazzy music.

4.2.2 COMPOSING SCENARIO

A couple of non-musicians, Aliaksei and Silvia, read an online post promoting an installation that opens a new possibility of composing music by means of movements in a room. Without hesitation, they decide to give it a try. Once in the room, they enjoy trying all possible combinations of distance and speed to produce changes in the music. Once acquainted with the functionality of the system, they focus on creating music with specific emotional character. They particularly enjoy the serene melody that seems to recur when they stand close, so they purposely stay in contact to generate this tune.

4.2.3 ACTING SCENARIO

Once instructed on the functionalities of the system, Maria and Zeno, a dyad of fervent theatre actors, realised that they can act out a drama while creating a soundtrack at the same time. After some consultation, they decide to perform a specific storyline: a tragic event with a happy ending. Once in the room, they initially spend some time exploring the interaction dimensions to ensure that the system correctly responds to their movements. Then, they start performing. Initially, they stand far from each other; then they eventually move to the opposite corners of the room. While maintaining the highest possible distance from each other, they alternate running with walking slowly. This variability results in an alternation between angry and sad music. After this tense and tragic stage, Maria gradually approaches Zeno. As she does so, the music assumes an increasingly cheerful character, and they eventually spend some minutes holding each other closely while a romantic musical composition plays.

4.3 TECHNOLOGY

The Music Room is composed of two main technological blocks: a tracking system and Robin, the algorithmic composer. The process of generating music from user movements involves three steps (Figure 4.2):

1. A video analysis tool detects the moving objects captured by a camera installed in the room in order to extract *proxemic cues* from the participants' behaviour.
2. These values are communicated to Robin that, in turn, transform the proxemic information into the emotional cues of *valence and arousal*.
3. Valence and arousal are transformed into combinations of *musical factors*, which determine the change produced in the generated music.

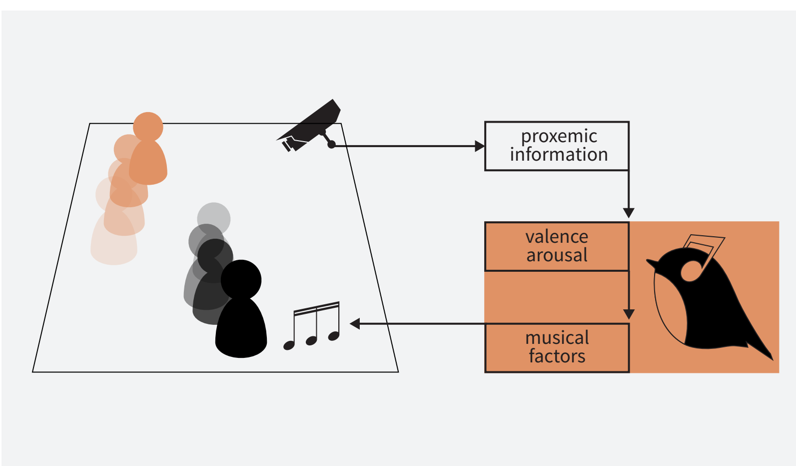


Figure 4.2. The architecture of The Music Room. On the left, the position and speed of the dyad is tracked with a camera that sends this information to Robin. Here, the proxemic information is interpreted in combinations of valence and arousal that provoke a number of changes in the musical factors.

4.3.1 VISUAL TRACKING SYSTEM

To make the installation more user-friendly and to minimise its intrusive qualities, the movements of the dyad are tracked with computer vision techniques rather than wearable sensors. Specifically, computer vision research was applied to interpret proxemic cues (KaewTraKulPong & Bowden, 2002; Rota, Conci and Sebe, 2012). Proxemics is the study of space in interpersonal interactions (Hall, 1973). It reflects the distance at which people are comfortable when talking to each other in a specific setting, thus providing important cues to infer the nature of social interactions and to understand such interactions. Different distances indeed mirror different kinds of relationships.

The motion of the participants was recorded through a downwards-looking bird's-eye-view video camera installed on the ceiling of the room. This configuration of the camera allowed the minimisation of the risk of occlusions, which is intrinsic to any motion tracking application, thus limiting the occurrences of false and missed detections. The detection of the moving subjects was implemented by applying a standard background subtraction algorithm (KaewTraKulPong & Bowden, 2001). The obtained foreground information was then processed by the CamShift tracking algorithm (Bradski, 1998). The choice of CamShift was primarily influenced by the necessity of keeping up with real-time constraints by reducing computational burden.

The position of the participants returned by the tracking algorithm was progressively updated over time, and the extracted proxemic cues were supplied to the system, for the purpose of providing information about the level of intimacy between them, which would in turn inform the music. Figure 4.3 displays a view of the room as seen by the camera. Two different instances of the interaction are shown on the left, while the output of the detection and the tracking module are portrayed on the right. In particular, the algorithm's capability of successfully managing partial occlusions, occurring when the two subjects approach, can be observed in Figure 4.3.d.

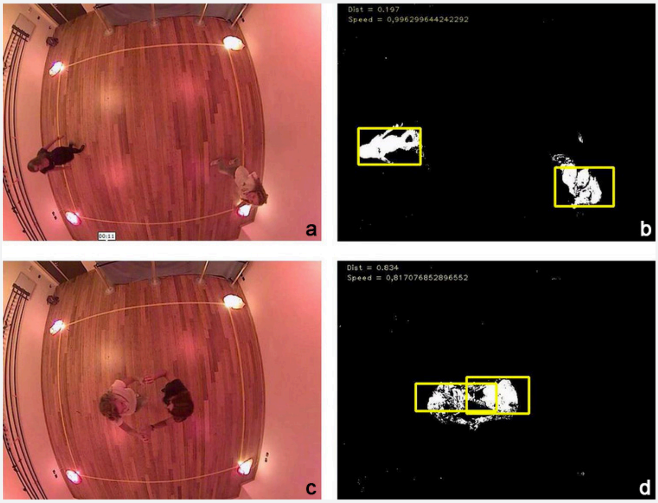


Figure 4.3. Two views of The Music Room as recorded from the camera mounted in the ceiling. On the right, the visual output of the motion tracking algorithm. The algorithm correctly detects the position of the two visitors even when their bodies partially overall (c-d).

4.3.2 ROBIN

Distance and speed are used by dyads to communicate the emotions they intend to communicate. Low distance is mapped with positive emotions and high distance with negative emotions; high speed with intense emotion and low speed with mild emotions. By matching the values of speed and distance to emotion, Robin adapts the musical flow in real time.

As previously discussed in Section 3.3, the emotional character of the music is continuously adjusted in real time, modifying seven musical factors: mode, tempo, pitch contour, pitch register, theme recurrence, sound level, and consonance. In addition, for the purpose of increasing the aesthetics of the composed music, different musical instruments were associated with different emotional conditions:

The *piano* is constantly present in all conditions;

A *violin* harmonises the piano voice when dyads are particularly close;

A *trombone* harmonises the piano voice when dyads are on the opposite sides of the room.

Figure 4.4. Mapping between musical factors and proxemic elements.

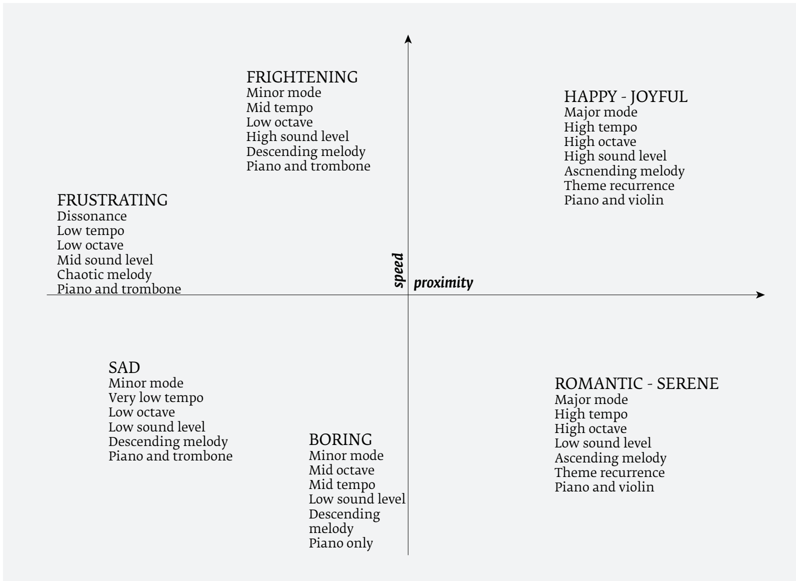


Figure 4.4 schematises how these parameters combine to communicate different emotions according to the two proxemic cues manipulated, distance and speed. Further details on this mapping can be found in Section 3.3.6. Robin is implemented in SuperCollider and it communicates via OSC with the tracking system. The output of the SuperCollider patch was a score in MIDI format. This score can be processed by any Digital Audio Workstation (e.g. Logic Pro, Ableton, Reason), which then transforms the MIDI flow into music.

4.3.3 PROTOTYPING

The architecture of Robin was first tested and fine-tuned with a low-fidelity prototype of the system. Initially, three short videos showing two individuals moving in a space were recorded. The videos were then fed into a preliminary version of the system, which extracted the proxemic cues and generated music following the proposed mapping. For each video, the generated music was then attached as an audio track. Without providing any contextual information, the videos were discussed by five HCI researchers in a design workshop. All of the participants reported the music as being somehow aligned with the movements of the couple and, in particular, their speed seemed to match the music intensity. Two people realised that their distance also had an effect on the music. Interestingly, participants also reported rules that did not actually apply. For instance, one noted: "when they move their hands in circle [sic] the music repeats". Another participant of the workshop reported: "note pitches are determined by the distance of their hands from the ground". This evaluation allowed testing of the quality of the architecture and evaluating the two scenarios with evolving prototypes. Finally, as the architecture developed into a stable system, the evaluation moved into the laboratory.

After the first exhibition of the installation, Robin was modified to accommodate visitors' preferences. Interviews conducted after the session (see Section 5.2.4) revealed that several participants complained about the latency between the user input and the musical response. This latency was due to some issues related to the tracking system as well as to a specific choice of avoiding sudden changes in music. This choice was made to preserve the phraseological structure of the music even in the case of rapid changes in the emotional input. For this purpose, the successive musical phrase was computed at the last beat of the playing bar, which was fixed at 4/4. This resulted in an approximately 4-second delay in the worst-case scenario, occurring when the current bar was at its first beat and playing at 60 BPM. In order to reduce this latency, while still preserving musical coherence, at every quarter beat, a new input from the user was checked. If it ranked above or below a specific threshold, a new bar started playing immediately. This solution reduced the latency time to 1 second in the worst-case scenario.

4.4 CONTEXT

This section details the context in which The Music Room was exhibited at three different venues.

4.4.1 FIRST EXHIBITION

The Music Room was first exhibited during the 2012 convocation of the EU Researchers' Night, which took place on September 28th in the city centre of Trento (Italy). This Europe-wide event involved 300 venues where academic and business research results were publicly showcased. In Trento, the event lasted from 5 PM to 2 AM, hosting almost 90 demonstrations and installations, which attracted a very heterogeneous audience. The Music Room was hosted at the Department of Humanities of the University of Trento in a 25 m^2 classroom, which had been previously emptied of all furniture. Some minor adjustments were made to the room: to make the environment more pleasant, the walls were decorated with musical patterns. The room was originally illuminated by some neon lights whose serious character did not match the intended mood of the installation. To this end, we covered them with orange veils to reduce the intensity of light and to provide a more enjoyable atmosphere. We wanted, indeed, to keep the room as dark as possible, for the purposes of fostering intimacy. However, we had to find a compromise solution, as the tracking system required a minimum level of lighting to function properly.

Once the room was set up we tested the tracking system. The test raised two issues: (i) the camera could not properly track the position of their position when they were standing close to the walls; (ii) the neon lights did not suffice enough light to correctly track visitor's position. The first issue was addressed by restricting the performance area, which was delimited by a sticky tape. The second issue was addressed by adding four light bulbs, which were also covered with orange veils (Figure 4.5).

4.4.2 SECOND EXHIBITION

The second exhibition of The Music Room took place on March 23rd 2013, on the occasion of the 5th edition of the ICT (Information and Communication Technologies) Days held at the Science Museum of Trento. The old, storied building located in the city centre provided an ideal setting for the installation. The 30 m^2 room that hosted the installation was once again emptied of all furniture. A group of students and researchers volunteered to decorate the walls with musical patterns (Figure 4.6). The event lasted for 8 hours (from 2 PM to 6 PM and from 9 PM to 1 AM) and the building hosted several others exhibitions. In particular, late in the evening, a disco-music concert was hosted on the ground floor of the museum. The event, and particularly the concert, attracted a significant number of young people. Many of them visited the floor where The Music Room was

Figure 4.5. A close-up of the orange veils covering the light bulbs in the floor of The Music Room at its first exhibition at the Researcher's Night 2012.

Figure 4.6. Decorating The Music Room at its second exhibition at ICT Days 2013.

also hosted and eventually participated in the installation.

4.4.3 THIRD EXHIBITION

The third exhibition of The Music Room was hosted at MART, the Art Museum of Rovereto and Trento (Italy), over the course of August 12^{th}-26^{th}, 2014. The installation was located in the educational area of the museum, which was open only by invitation. The space was a 24 m^2 room, typically employed as an art gallery. By requirements, the room was entirely emptied from of furniture except for a couple of immovable elements: a painting hung on one wall and a sink was present. Four speakers were positioned in the corners of the room. The tracking camera was mounted on one of the beams of the ceiling and a second camera was also installed in the room to facilitate the video analysis. Six light spots ensured that participants' positions could be properly tracked. Finally, the windows were covered with long blank sheets to prevent external lights from interfering with the tracking system (Figure 4.7).

Figure 4.7. The Music Room as exhibited at MART.

4.5 CONCLUSION

This chapter presented the design of The Music Room, an interactive installation that allows musically untrained visitors to compose an original music by moving throughout a room. The conceptual design and the prototyping stage of the system were also presented. Three possible interaction scenarios were envisioned. The next chapter focuses on understanding whether and how visitors' experience matched our design expectations with a field study and a controlled study.

Interface Evaluation: The Music Room

THIS CHAPTER DISCUSSES THE EVALUATION OF THE MUSIC ROOM, WITH A FOCUS ON THE EXPERIENCE OF THE AUDIENCE. A QUANTITATIVE STUDY PERFORMED WITH THE GENERAL PUBLIC AND A QUALITATIVE STUDY PERFORMED WITH INVITED COMMENTATORS ARE DETAILED, CONTRIBUTING TO THE UNDERSTANDING OF VISITORS' ENGAGEMENT WITH INTERACTIVE INSTALLATIONS.

5

5.1 INTRODUCTION

This chapter presents the evaluation of The Music Room as conducted in two studies. The first study focused on understanding visitors' behaviours with The Music Room analysing data collected from two field studies in which the installation was publicly exhibited (Section 5.2). An integration of log-data, video analysis and subjective evaluation was performed. The analysis offered insights to understand the experiences of the visitors who engaged with the installation in a number of different ways, at times appropriating the original design idea. Reconsidering design assumptions compared to behavioural data, a number of unexpected behaviours were noted. In particular, the evaluation suggested a clear distinction between visitors who actively engaged in composing music and those who passively moved to the music. Both behaviours were often manifested in the form of dancing.

The second study was arranged to clarify this behavioural ambivalence and to provide further information about the variability of users' experiences (Section 5.3). This time, a qualitative approach was adopted. In-depth feedback was collected as to visualise the audience experience from individual perspectives rather than to search for average group behaviours. Visitors' interpretations contributed to understanding how the general public could experience the installation. In particular, it was disclosed that (i) the installation offered a number of non-ordinary experiences, and (ii) visitor engagement was sustained by a variety of experiences, either connected to an exploration of the functionality of the system or to an engrossing involvement with the installation. Providing visitors with minimal information on system functionality increased the creative engagement of the visitors, but their experiences mirrored to a lesser extent that engagement that had originally been envisioned.

5.2 FIELD EVALUATIONS

The two field studies were performed in Trento on the occasions of the 2012 convocation of the EU Researchers' Night and on the occasion of the 5^{th} edition of the ICT Days held at the Science Museum of Trento (refer to Section 4.4.1 and 4.4.2).

For both occasions, before entering the room, visitors were informed that they could direct the music, which was generated by a computer, through their own movements. Their distance from other visitors would influence the pleasantness of the music while their speed would change the intensity (i.e. volume and tempo). After this brief explanation, a researcher invited participants to sign a consent form notifying them that their session would be videotaped. Then, the researcher left the room and closed the door. Initially, people were free to experience the installation for as long as they wished. However, for both events, as the waiting queues dramatically increased over time, several people were invited to leave the room after they had participated in the installation for a few minutes.

Once the members of the dyad had left the room, two researchers addressed them with a few questions. In addition, on the occasion of the second exhibition only, they were given a card containing the URL and the QR-code to an online questionnaire, together with a personal code. With this code, once they had completed the questionnaire, they could download the music they created during the event. In addition to the camera installed on the ceiling of the room, which was used to track participants' movements, another camera was mounted in the room to videotape their performance from an additional perspective. This further point of view allowed the researchers to gain a better understanding of the behaviours and engagement of the participants.

The results presented in the following sections are based on the integration of online observations, interviews, questionnaires, log-data and video analysis.

5.2.1 FIELD OBSERVATIONS

Posters placed across the venues (Figure 5.1) advertised the installation and contributed to attracting visitors. For both exhibitions, The Music Room was constantly busy from the opening to the very end: 87 and 85 dyads participated, for a total of 344 visitors of all ages who used the installation. Individual visits lasted on average 5 minutes each (from a minimum of 1 min 30 s to a maximum of 10 min). At both venues, the installations quickly gathered increasing

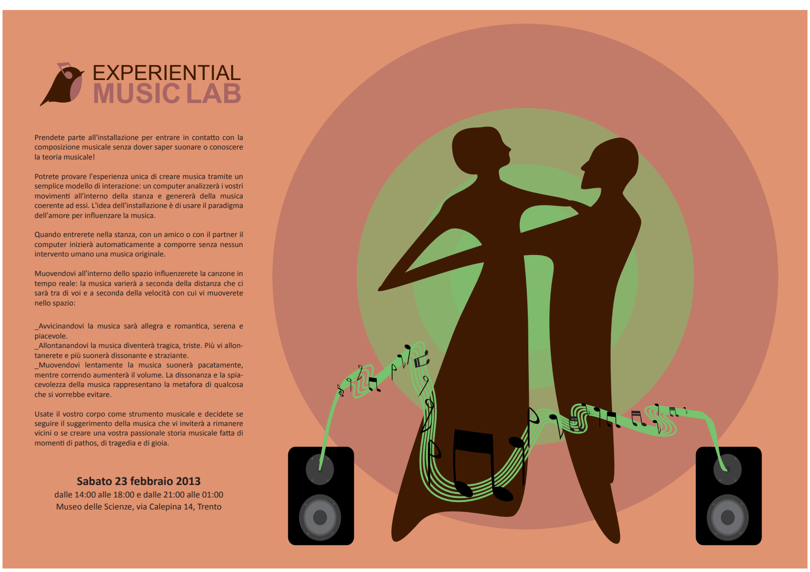


Figure 5.1: One of the posters that contributed to attracting visitors' attention.

success, as witnessed by the long queues of visitors. However, the smiling faces of the visitors leaving the room, and a video demonstration running nearby, appeared to be an attraction to many people. Indeed, despite discouraging several visitors, the long queue also caught the attention of several passers-by who eventually ended up joining the queue.

During the first exhibition, we were caught unprepared for such a success, but a team of six generous and passionate colleagues and friends volunteered to assist us. As the installation became busy and visitors started lining up to try it, they were invited to write their names on a waiting list. They were provided with an estimate of their waiting time, so that in the meantime they could continue their visit, to come back to the room once their turn was actually approaching. Despite of this, at 2AM we had to open the room to the final curious visitors who were able to experience a quick and unpredictable group music creation.

Both nights were extremely intense and passed by with a hectic euphoria and almost no technical incidents, but it taught us several important lessons about the logistics of our research settings. During the second exhibition we organised shifts involving a team of 13 researchers who worked hard throughout the exhibition period.

5.2.2 VIDEO ANALYSIS

The video analysis is based on samples of videos (N=50) collected from both exhibitions equally. The investigation aimed at probing into the most recurring behaviours exhibited by the visitors. Categories of behaviours were isolated and fine-tuned with the help of two researchers who independently viewed the video footages of each session several times. At the end of this stage of analysis, two categories of behaviours had been identified: *acting* and *inter-acting*. *Acting* refers to exhibiting independent behaviours: the dyad members did not physically interfere with the behaviour of each other. *Inter-acting* occurred when the members moved together in such a way that one person's movements directly affected the other person's movements, occasionally with direct physical contact between the two bodies. These two types of behaviour normally coexisted as visitors naturally alternated from one to the other.

Results are summarised in Table 5.1, showing the percentage of dyads that performed that particular behaviour and the percentage of time spent performing that particular behaviour, considering all of the dyads. Alongside, the inter-rater agreement (i.e. Cohen's kappa) for each behavioural theme is provided. The kappa values range from 0.343 to 0.824 with an average of 0.601. Following the Altman (1990) interpretation of the kappa vale, the values then ranged from fair (0.21-0.40) to almost perfect (0.81-1.00), thus demonstrating, on average, good reliability of the data.

The four most common *acting* behaviours were *walking*, *running*, *dancing*, and *standing still*; all of these actions had a direct effect on the music played in the room. A considerable number of participants experimented with other playing behaviour, such as *jumping*, *lying on the floor* (and at times *spinning* or *rolling*), *twisting*, *bowing*, or *stamping the ground*, or *mimicking a love declaration*. These behaviours did not have a direct influence on the music, and they mostly occurred at the apex of the experience, when people looked particularly engaged. *Inter-acting* behaviours ranged from a playful *run and chase game*, or the enactment of a *fight*, to more intimate experiences such as *couple dancing*, *hugging*, or the more vigorous *pirouetting*. A number of visitors also laid down and *crawled*, as if attempting to *confuse* the system by hiding from the camera view.

Table 5.1: List of the most common behaviours exhibited with associated values of Kappa.

ACTING	% OF TIME	% OF DYADS	KAPPA
Walking	36	96	0.613
Individual dancing	11	56	0.690
Running	10	82	0.612
Standing still	8	100	0.522
Lie down	3	7	0.824
Jumping	2	47	0.599
Twisting	4	28	0.343
Bowing	0.5	16	
Stamping the ground	0.4	8	
Mimicking declaration of love	0.2	2	
INTER-ACTING	% OF TIME	% OF DYADS	KAPPA
Couple dancing	17	76	0.762
Run and chase game	6	64	0.641
Pirouetting	2	32	0.612
Intimate behaviours	2	32	0.409
Lift	1	16	0.591
Fight	0.5	28	0.595

This analysis offered a detailed understanding of the behaviours exhibited by the visitors to this installation. Their engagement with the installation undeniably emerged from a number of performed actions; the high occurrence of several of these actions, however, indicated that they did not simply experience the installation as a musical controller.

5.2.3 LOG DATA ANALYSIS

During the second exhibition, the log data describing dyads' (N=63) position and speed was collected and stored to be later analysed.

The first analysis examined possible common interaction trajectories (i.e. performance patterns) by analysing the movements of each dyad during the entire session. We were expecting a particular trend to be exhibited by a considerable number of dyads. Specifically, in a temporal order:

1. Once in the room, visitors would try the most extreme combinations of musical output (e.g. standing still close and far, and running). This behaviour would result in a substantial variability of both distance and speed.

2. Once acquainted with the system, they would engage in dancing, chasing play or performing intimate behaviours. Distance and speed would vary considerably at this stage.

3. Before exiting the room, visitors would spend a short period disengaging from the activity. This would correspond to a gradual reduction of speed and stabilised distance.

To test this prediction, for each dyad, information on speed and distance was plotted and visually inspected with the help of two researchers. Results contradicted our prediction: the expected pattern did not emerge from the visual inspection of the plots, and largely varied among dyads. In some cases, proxemic cues varied to a very limited extent, suggesting that visitors continued to perform the same behaviours. By contrast, other dyads continuously changed their speed and distance. Figure 5.2 shows the distance between the members of a dyad and their average speed as they varied during a session. This data was representative of a typical data set with respect to the continuous variability of the two dimensions.

The second analysis focused on understanding whether the average values of distance and speed, collected during each session, could provide interesting insights about visitors' experiences. In particular, the combination of the means of distance and speed could provide information about visitors' behaviours (Table V.2). For instance, intimate slow dancing could be represented by low means for both distance and speed, and running by high means for both factors. By contrast, divergent combinations of the two means (i.e. high-speed with low distance, and low-speed with high distance) could either refer to fast movements performed while being linked together (pirouetting) or to individual behaviours performed at a high distance. For the sake of convenience, the values of the means of distance and speed were divided into three categories with the 33^{rd} and 66^{th} percentiles. Table 5.2 maps combinations of the means for distance and speed with characteristic behaviours.

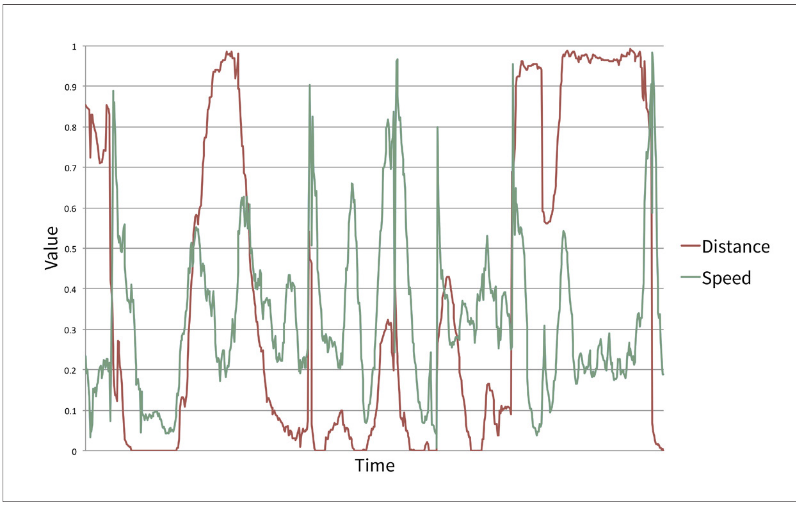


Figure 5.2: Speed and distance variability during a single session. Both distance and speed continuously change, thus suggesting that the dyad members are trying different combinations of musical outputs.

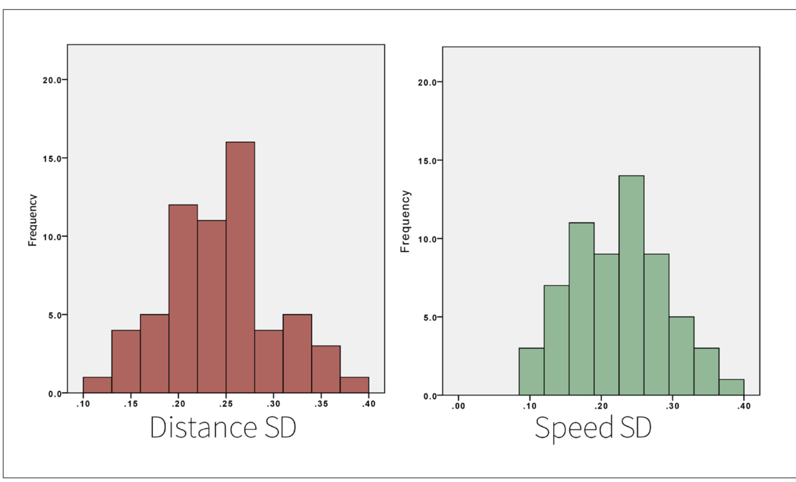


Figure 5.3: The histograms of the standard deviations of distance and speed show a Gaussian trend, suggesting that performed behaviours widely varied between and within dyads.

Table 5.2: Predicting behaviours by means of means and SD of distance and speed.

	LOW DISTANCE	MEDIUM DISTANCE	HIGH DISTANCE
LOW SPEED	Intimate behaviours Talking / discussing Collaborative dancing Romantic dancing	Individual and collaborative dancing	Individual behaviours
MEDIUM SPEED	Dancing Walking together	Walking	Individual behaviours
HIGH SPEED	Performing pirouettes Fighting	Pursuit	Running

Results revealed that the means for both distance and speed greatly varied among dyads. The nine combinations had a very similar incidence, bearing witness to the great variety of behaviours manifested in the room. The most common combination was high speed with high distance (14.3%), associated with running or playing.

To examine this variability in more detail, the standard deviations (SDs) of distance and speed were analysed. Low SDs for both variables would suggest that visitors did not change their behaviours by any significant amount. Rather than exploring all of the interaction possibilities, they preferred to adopt a more passive

behaviour. On the other hand, high SDs for both variables would suggest that visitors spent a significant percentage of their time experimenting with different combinations of speed and distance. The frequencies of the two SDs showed that both variables exhibited a Gaussian distribution among the dataset of all of the dyads, confirming that performed behaviours widely varied between and within dyads (Figure 5.3).
Log data were also analysed to evaluate potential differences among genre distribution. The means and SD values of distance and speed were entered as dependent variables into a MANOVA analysis with genre distribution as the between-subjects factor. Results indicated that the dyad composition exerted a significant effect on average distance ($F(2,60) = 3.47$, $p < .1$, $\eta p2 = .1$) and speed standard deviation ($F(2,60) = 2.48$, $p < .1$, $\eta p2 = .07$). The effects were due to mixed-gender dyads, which interacted at a closer distance and varied their speed on a less frequent basis than male-only dyads. This result can be explained with reference to the stereotype of social acceptance of physical proximity in dancing situations.

5.2.4 INTERVIEWS

At the first exhibition, 63 dyads were asked three questions and encouraged to express any further comments if they wished to do so.

1. The first question probed visitors' general experience, which was described by almost all visitors with flattering words (e.g. *cool*, *interesting*, *unique*, *intimate*, *pleasant* and *relaxing*). The only two visitors who did not enjoy the experience complained about a lack of interactivity: they were expecting a more direct manipulation of the artistic artefact. Twelve dyads reported that they were particularly impressed by the quality of the music.

2. The second question invited visitors to elaborate on the negative aspects of the experience, and suggest possible improvements. Once again, most of the people were notably positive. The only concern, which was shared among 14 dyads, addressed a delay between their movements and the music reaction. This issue was solved for the second exhibition improving the response delay of Robin to distance and speed changes (see Section 4.3.3).

3. The last question investigated the extent to which visitors perceived being in control of the music. This question highlighted an important dichotomy. Nearly half of the interviewed

participants reported that they felt as if they were actively controlling the music. The other half declared that they were mainly following the music, only having the impression of playing an active role in a few situations. For instance, six dyads reported that they had initially spent some time controlling the music, but then had simply forgot about the instructions and subsequently followed the music.

In the second exhibition, 77 dyads were interviewed. This time, we were mostly interested in understanding whether the ambivalence between controlling vs. following the music had been reduced by the technical intervention on Robin (Section 4.3.3). Results showed that this was just partially the case, as only 58% of the interviewees reported that they were controlling the music; the remaining visitors felt like they were following the music (15%) or experienced both feelings (27%).

5.2.5 QUESTIONNAIRES

For the second exhibition, a total of 57 questionnaires were collected from 32 female and 25 male respondents, 26% of whom reported being capable of playing an instrument. The number of respondents was particularly high: with respect to the total number of visitors, 34% of them responded to the questionnaire. The last page of the questionnaire showed a textbox where participants could enter the code they were given at the end of the session and download the song they had created in the room. They were also invited to visit our website to find information on our work, and to leave further comments on our Facebook page.

The seven questions, with means and SDs, are listed in Table 5.3. The music was generally appreciated (3.93), although some visitors would have preferred other musical genres. The most negative response regarded the number of available movements used to influence the music (2.77).

Table 5.3: Results of the online questionnaire.

ITEM	MEAN (SD)
"I did not like the music inside the room" (reversed)	3.93 (SD=.98)
"I enjoyed the installation"	4.18 (SD=.98)
"It was a stimulating creative experience"	3.77(SD=1.17)
"I will recommend my friends to try this installation"	4.33 (SD=1)
"The music followed my and my partner's movements"	3.15(SD=.97)
"The number of available movements to influence music were too few"	2.77(SD=1.28)
"I would have preferred other musical genres" (reversed)	3.02 (SD=1.51)

The data were analysed by means of a principal component factor analysis with Varimax rotation (Kaiser Normalisation). Two components with an eigenvalue of greater than 1.0 were found. The components can be thought of as representing the general engagement with the experience (Component 1) and possible changes on the musical interaction (Component 2). The components loading are shown in Table 5.4.

A parametric test of correlation was then performed between expertise and the two components. There was a significant negative correlation between expertise and Component 1 ($r = -.316$, $N=57$, $p<.05$, one-tailed), thus suggesting that non-musicians had a more engaging experience.

Table 5.4: The components found by the principal component analysis, and variables that load on them.

COMPONENT 1	
"I enjoyed the installation"	.876
"I will recommend my friends to try this installation"	.801
"It was a stimulating creative experience"	.647
"I did not like the music inside the room"	.640
"The music followed my and my partner's movements"	.577
COMPONENT 2	
"I would have preferred other musical genres"	.876
"The number of available movements to influence music were too few"	.801

5.2.6 SYSTEM QUALITY AND RELIABILITY

In addition to examining the experience of the visitors, we were keen to analyse the quality of the system, i.e. the accuracy of the response of the system to the movements of the users. In particular, we investigated (i) how precisely dyads' positions were tracked, and (ii) how promptly Robin adapted its musical output in response to visitors' movements.

To obtain information about the accuracy of the tracking system, during both exhibitions, two researchers sat behind the control desk, observing the reaction of the system to the movements of the visitors. The visual tracking algorithm tracked dyads' positions fairly accurately. Occasionally, when the members of the dyads were standing in close proximity for an extended period, the system was observed to lose track of one of them. In most of the cases, the system recovered from this error very quickly. However, in very few occasions, we were forced to reset the system, thus losing track of the position of the dyads for approximately 10 seconds.

With respect to Robin, interviews revealed that most visitors were impressed by the quality of the music, which was often described as barely distinguishable from that produced by a human musician. However, during the first exhibition, a brief latency between user movements and the generated music was reported. This latency was a consequence of a precise design choice: we intentionally decided to avoid sudden changes in music in order to preserve the phraseological structure of music, even in the case of a sudden change in the emotional input. Several visitors, however, commented on being annoyed by this latency, as they were expecting the music to change instantaneously in response to their movements. In order to fulfil this request, for the second edition of The Music Room, we modified the algorithm to reduce the latency (Section 4.3.3).

5.2.7 DISCUSSION

Results collected from field observations, interviews, video analysis and questionnaires confirmed that a large percentage of visitors deeply enjoyed The Music Room. This enjoyment was due to a full range of different pleasurable behaviours, which were identified by log data and video analysis. However, both techniques failed to assess the reasons that motivated this different behaviours, and to identify the factors that accounted for the diversity of experiences. Following reflections from related studies, this difference might be attributable to the diversity of visitors' "interpretation, understanding, attitudes, personality and expectations of computer culture" (Höök et al., 2003).

This study also evidenced that the three scenarios originally envisioned in the conceptual design phase (*playing, dancing* and *acting*

- Section 4.2) occurred with differing incidence. The most common scenario was *dancing* (performed by 76% of the dyads), probably because the synergy between music, movements and emotions is often associated with dancing. The *acting* scenario occurred only occasionally: during the interviews, only two dyads reported that they had been pretending to act in a theatre, and two other dyads mimicked love declarations, as observed by the video analysis. These two scenarios were reasonably easy to detect via video analysis. By contrast, the *composing* scenario could not be easily assessed via this technique, which failed to assess whether dyads ran, jumped, danced and walked to consciously influence the music. For instance, in order to make the music more tragic, dyad members could have simply walked away from each other or performed the same action while jumping. As a further example, to create a serene musical output, they could have simply stood in close proximity or danced intimately.

The interviews provided us with better insights on this topic. The question concerning the level of active involvement in the music process indeed indicated that approximately half of the dyads purposely tried to control the music. In addition to the original scenarios, the field studies disclosed a number of behaviours that had not been envisioned during the design stage. In particular, behaviours expressing delight and excitement occurred with a high incidence: *pirouetting*, *twisting,* and enactment of a *fight* were each performed by one out of three dyads. Furthermore, the analysis of the videos and the log data revealed that one third of the dyads engaged in intimate behaviours such as *romantic dancing*, *kissing*, *hugging* and *lifting*, confirming the potential of The Music Room to facilitate intimate experiences. In these cases, the association between gestures, music and emotions possibly recalled memories from romantic movies or personal histories.

In the light of the initial objective of the installation, participants' ambivalent perception of the degree of control over the musical output was probably the most unexpected result. The evaluation also revealed that, rather than simply focusing on making music, most of the visitors spent a notable amount of time performing actions that were not directly connected with music generation. Given the interactive dimensions at participants' disposal (i.e. distance and relative speed), the only gestures that would have a direct influence on the music were *walking, running* and *standing*. However, video analysis revealed that these gestures globally accounted for approximately just about half of the recorded time. Visitors appropriated the installation giving it their personal interpretation. This observation suggested that, in this design area, the actual status of the work could be defined and fully understood only when submitted to audience verdict. In fact, The Music Room, ideated as a novel musical controller, showed its status only when analysed with the interactions and the behaviours of the users taken into account.

Finally, it is worth noting that the techniques adopted in this study successfully accounted the most recurring behaviours exhibited in the room, but they generally failed to explain the motivations that produced them. In the next paragraph, we weigh the costs and benefits of the different data collection techniques:

Field observations provided an initial understating of the engagement of the visitors. Following the framework proposed by Edmonds (2010) for understanding engagement with interactive art, this methodology contributed to clarifying what the *attractors* of the installation were, i.e. the poster and videos placed all over the venue, the long queue and the smiling faces of people leaving the room.

Video analysis proved crucial for assessing visitors' behaviours in the interactive installations, allowing a precise understanding of the most common behaviours exhibited as well as individual performances. However, performing accurate video analysis is a time-consuming method that requires engaging several researchers at time. In addition, this information alone does not suffice to infer the driving motivations for visitors to engage in their selected activities.

Log data analysis is a minimally time-consuming technique that might prove useful to gain a general knowledge of the variability of visitors' experiences with an interactive installation. In this case, it allowed us to acquire a better understanding of the variability of their experience, thus corroborating the thesis that The Music Room can foster diverse behaviours.

Interviews collected feedbacks that helped to clarify visitors' behaviours. Collecting impromptu comments proved an effective resource to understand the first impressions of the visitors, i.e. the factors that most significantly caught their attention and *sustained* their engagement (Edmonds, 2010). Among the limitations of this approach, we cite the difficulty of conducting such a study during public exhibitions.

Questionnaires are as good as the questions they contain. Indeed, if properly administered, this technique can reveal interesting quantitative insights about visitors' experiences at a relatively low cost. However, this methodology can be effective only if a sufficient number of entries are collected. Thus, a designer who is willing to exploit this technique should carefully ponder when and by which means to collect questionnaires. To acquire as many completed questionnaires as possible, one solution would be to administer them immediately after the experience. An alternative solution is to administer questionnaires online. By doing so, the percentage of visitors that make the effort to go online at a later time to complete the questionnaires would itself be an indicator of the participants' appreciation of this installation.

5.3 CONTROLLED EVALUATION

A new exhibition of The Music Room was arranged for the purpose of clarifying some of the issues that had remained unsolved following the field studies. In addition, we aimed to understand whether and how the issued narratives of use influence visitors' experience. This time, the installation was tested at MART, the Art Museum of Rovereto and Trento (Italy) (Section 4.4.3) with invited visitors and evaluated though in-depth interviews, focusing on idiosyncratic interpretations rather than group reactions that are hypotesised to hold for everyone (Höök et al., 2003).

With this end in view, 13 commentators were invited by e-mail to participate in the study and to bring along a person of their choice they felt comfortable testing the installation with (from now on, we will refer to *commentators* regardless of whether they are the individuals we invited or their partners). One commentator could not find a partner so he was matched with an art expert who was available. To gain multifaceted feedback, commentators both inside (Šimbelis et al., 2014) and outside (Gaver, 2007) the communities of interest were involved: music, contemporary art, dance, theatre, neuroscience, computer science, psychology, sociology and neuroscience field were represented. Participants were selected not only on the basis of their professional profile, but also for their artistic sensibilities.

The conversation with the visitors revealed that two couples that had been originally assigned to the no-information group actually did already know something about the installation, so they were re-assigned to the full-information group and given complete verbal information.

In total, 26 people (17 females), aged between 21 and 58 years used the installation. Table 5.5 lists the characteristics of the commentators. Their age, gender, expertise area, and artistic interests are also featured, along with the kind of relation linking them to the partners they had brought along for the study, and the amount of information they were issued. To indicate a specific commentator, in the remainder of this paper, we will use the associated number followed by the letter "*i*" (invited) or "*p*" (partner).

Table 5.5: The commentators that were invited to try the installation. R column: relation between the invited commentators and their partners (C=colleagues, F=friends, P=partners, S=strangers). I column: amount of information received (0=none, 1=partial, 2=full).

#	INVITED	PARTNER	R	I
1	31 M - wikipedian in residence, computer scientist	29 F - museum employee and Jazz pianist	C	1
2	21 M - actor, director, amateur musician	21 F - studies at the conservatory of music	F	0
3	28 F - physician, student art & science	38 M - physician, studied music	P	1
4	41 M - copywriter, singer, dancer	45 F - trainer	P	2
5	57 F – piano teacher at the conservatory of music	58 F - civil servant, studies music	C	2
6	41 F - art historian	30 F - museum employee, studied music	C	1
7	29 M – robotic expert, amateur musician	41 F - art expert	S	2
8	59 F - prof. of clinical psychopathology	49 M - music therapist, musician	C	2
9	46 F - prof. of social psychology, had played the piano	39 M - prof. of social psychology, amateur drummer	P	0
10	42 F - art critic, curator, researcher	43 F - neuroscientist, had played the harp	F	1
11	52 M - dance teacher, choreographer, musician	26 F - psychologist, dancer, plays clarinet	C	1
12	33 F – social psychologist, plays the piano	33 M - graphic designer	F	0
13	30 F - behavioural neuroscientist, stage actor	30 F - educator, painter	F	0

5.3.1 PROCEDURE

At their arrival, the commentators were welcomed by two researchers, who asked them to sign a consent form, explaining that the session would be videotaped. They also had to specify their occupation, musical knowledge and artistic interests (if any), and the kind of relation that associated them with the partner they had brought along. They were also asked if they were already aware of how the installation worked. Everybody was informed that they could enjoy the installation for as long as they wished. In addition, the researchers specified that the objective of the study was to gather interpretations and reflections from commentators with specific expertise. Positive and negative feedback would prove equally valuable to us. Most importantly, they were informed that their behaviours in the room would not be judged in any way. Emphasising this point was crucial to making commentators feel comfortable and to fostering their trust and confidence. Finally, they were invited to feel free to do whatever they wished in the room.

Commentators were randomly provided with one of three levels of information. Four dyads were given complete verbal information: a researcher explained in detail the motivation of the study (i.e. to create an instrument to enable all users to experience music making), how the installation worked, and how they could

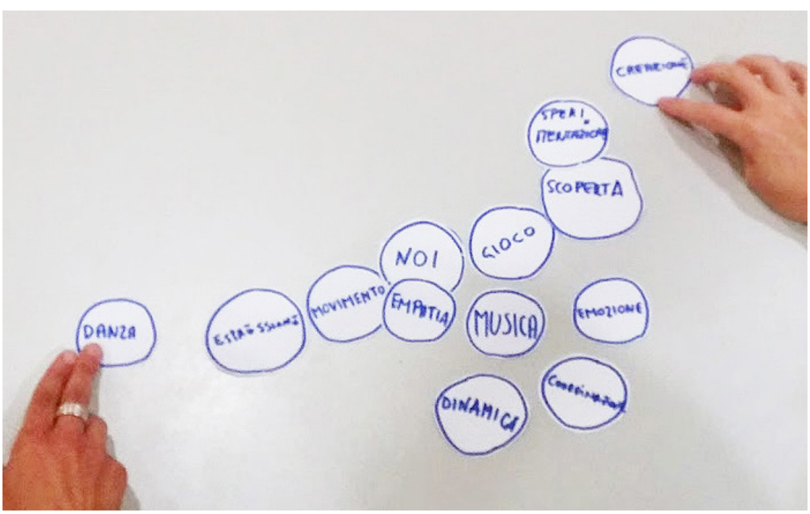


Figure 5.4: Concept mapping exercise. The commentators freely selected the concepts that better applied to their experience.

influence the music by means of their own movements. Five dyads only received ambiguous descriptive information on the mechanism underlying the installation. A sentence written on a piece of paper read: "*Freely move in the room, your **proximity** and **speed** will change the music*". The last four dyads did not receive any information at all: irrespective of their requests they were simply informed that The Music Room was an interactive installation about music.

After the dyads had left the room, they were invited to sit around a table with the two researchers and were offered water and drinks, as many of them had particularly long and intense sessions. Commentators were interviewed as a dyad. During the first stages of the interview, they were asked to freely report any thoughts, reflections and criticisms about the experience in their own terms, rather than being influenced by researchers' projections. In most cases, they kept talking for several minutes. Sometimes, instead, the researchers had to urge them on with general questions (e.g. "*Did you have any suggestive feeling?*") and direct questions (e.g. "*How did the music sound like when you were close together?*"). To avoid biasing the study, an external researcher, who had not taken part in the project, conducted the interviews. Before concluding the interview, the commentators were presented with a simplification of the concept mapping method for guiding evaluation (Trochim, 1989).

This method was proposed to the participants when we noticed that they were becoming less engaged in the interview. They were shown a number of hand-drawn circles (Figure 5.4): each of the circles contained a word that conveyed a concept potentially related to their experience. Participants were asked to choose the ones that best applied to their experience, and to arrange them in relation to each other (e.g. in a hierarchical or relational order). In addition, they were provided with empty circles, where they could add their own words if they wished to. The "new" words were added to the range of circles that would be shown to the successive dyads. At the end of the experiment, 13 circles had been added to the initial set of ten words, chosen by the researchers (i.e. *me*, *partner*, *game*, *room*, *movement*, *music*, *dance*, *empathy*, *emotion* and *creation*). During the concept mapping, no explicit question was asked, as the activity was not goal-structured or action-oriented.

5.3.2 DATA ANALYSIS

The data source of the investigation was composed of an integration of three evaluation methods: interviews, analysis of the concept mapping, and video analysis.

1. *Interview transcripts.* Interviews lasted from 21 to 38 minutes (average 31.2). Two researchers produced written transcripts of the 13 interviews with additional notes about contextual circumstances (e.g. behaviour in the room, relationship between the dyad, amount of information provided, personal background).

2. *Concept mapping.* On average, this method accounted for 34% of the interview time. In particular, it proved remarkably effective for those who were brief in the interview, reaching a peak of 70% for the dyad who engaged with the interview for the shortest time. All commentators narrated their experience by selecting and discussing the concepts in an order that mirrored their relevance with respect to the overall experience (e.g., "I would start from DISCOVERY. It was mainly about understanding the functionality of the system"). A list of the concepts selected by each commentator, as well as the transcript of the activity, was produced.

3. *Video analysis.* The 13 videos were analysed with the help of a second researcher, with the objective of gaining insights on the experience of the visitors and discovering singular behaviours.[1]

This analysis produced written annotations of (i) the experiences of each visitor, jotting down potential evolutions of their behaviours, and (ii) particular behaviours exhibited by the dyads (e.g. "[2i] they continuously interacted with the sink, trying to understand whether it influenced the music").

The methodology to interpret these data evolved with several iterations of inductive thematic analyses (Braun & Clarke, 2006). The interview and the concept mapping transcripts, as well as the video analysis annotations, were analysed with the help of a researcher. The most interesting quotes and notes with respect to the research goal of understanding audience experience were highlighted at this stage and associated with a code. Codes were iteratively analysed and clustered into themes following a semantic approach. For instance, the code *information* (reflections on the effect of the information provided in the study) and *interpretation* (verbal explanations of the system behaviour), which counted respectively 9 and 6 codes, were clustered into the theme *narrative of use*.

1 An extract of a video recorded in the room can be found at http://youtu.be/DrULqzx7p-Mo

Ultimately, the analysis revealed four main themes: *non-ordinary experience, modalities of engagement, interpreting narratives of use* and *idiosyncratic interpretation*.

5.3.3 NON-ORDINARY EXPERIENCE

Sixteen visitors noted that the peculiar character of the installation is that it offered a *non-ordinary experience*. Overall, it appears that The Music Room appealed to the satisfaction of self-actualization needs, one of the highest-level elements of human well-being (Maslow, 1943). Five dyads credited this feeling to being empowered to make music simply by means of their movements: "I felt god-almighty. Using my body to create a classical music was amazing!" [10i]. Some of them emphasised that they had been enabled to have control on music for the very first time in their life: "For the first time I could actually control music...and simply through my own movements." [4p]. One visitor even questioned the concept of authorship of the composition. Another commentator claimed that: "On one occasion the music was particularly sad and melancholic. Then, we just realised that it was up to us to make it happier by getting closer to each other...so let's change it!" [12p].

Video analysis revealed that dancing was one of the most common behaviours exhibited in the room, as 12 commentators spent some time participating in individual or couple dancing. Interestingly, three commentators reported that the installation offered a non-ordinary experience insofar as it subverted known rules of dancing. "Usually, when you dance it is your body that follows the music. It was cool that we were influencing the music and not the other way round" [7a]. A dyad of contemporary dancers put the emphasis on this point and explicitly stated they would use the installation in their work: "It is like a choreography, just the other way round. We move and the system crafts 'a musical dress' on top of our movements" [11i]. The subversion of the rules of dancing can prove helpful to people who feel uncomfortable when dancing: "I personally have issues with dancing. I simply cannot move to the music. But now that I was the one who was creating it, I just lost myself in the experience and danced" [4p].

5.3.4 Modalities of engagement: exploration vs. flow

The engagement with the installation was mainly sustained by two factors: *exploration* and *flow*. Most of the dyads (N=10) enjoyed the installation as an exploration of the functionality of the system. Their goal was to understand how it worked, how they could interact with it and how they could tweak it: “My main concern was to understand which options I had, what I could change” [9p]; “I wanted to put the system to the test, so I tried to hide and then to suddenly change rhythm and direction” [2i]; “I tried to make sense of how the system worked. Understanding what happens when we get close, when we move apart, when our movements are coordinated...” [6p].

Three dyads attributed the engagement to having experienced the typical conditions of flow (Csikszentmihalyi, 1991). “The emotional aspect was extremely intense. I definitely had fun, but it was not just about having fun. It was emotionally intense. I was literally swallowed up by the installation” [4p]. “The most beautiful aspect was that I completely focused on the here and now. In everyday life, it is almost impossible to achieve this state of mind” [10i], and “I enjoyed the room as an introspective experience. That was my objective” [4i].

Rational exploration and flow almost never overlapped: “It’s either about creating or having fun...understanding or letting it flow” [13i]. “In my opinion it was not about fun or emotions: I had another objective, I wanted to discover the trick. Had I found it, I would have enjoyed” [9p]. “You can either focus on the emotional or on the technical aspect. I think it makes more sense as an emotional experience rather than a musical instrument” [7i]. In some cases, these two conditions alternated during the experience: “At the beginning, the most important element was the emotional factor. It was associated with the music and the presence of the other person. Then, we started talking and planning the interaction. The emotional aspect retreated into the background, and we experimented” [7i]. “I was primarily interested in investigating how the system worked; if I were to try it again it would be less entertaining but more creative and emotional” [1p]. In one case we were confronted with the opposite scenario: a couple of friends started to purposely control the music, to eventually discover that their efforts were unsuccessful. Therefore, they stopped trying to control the music and simply had fun dancing and clowning about.

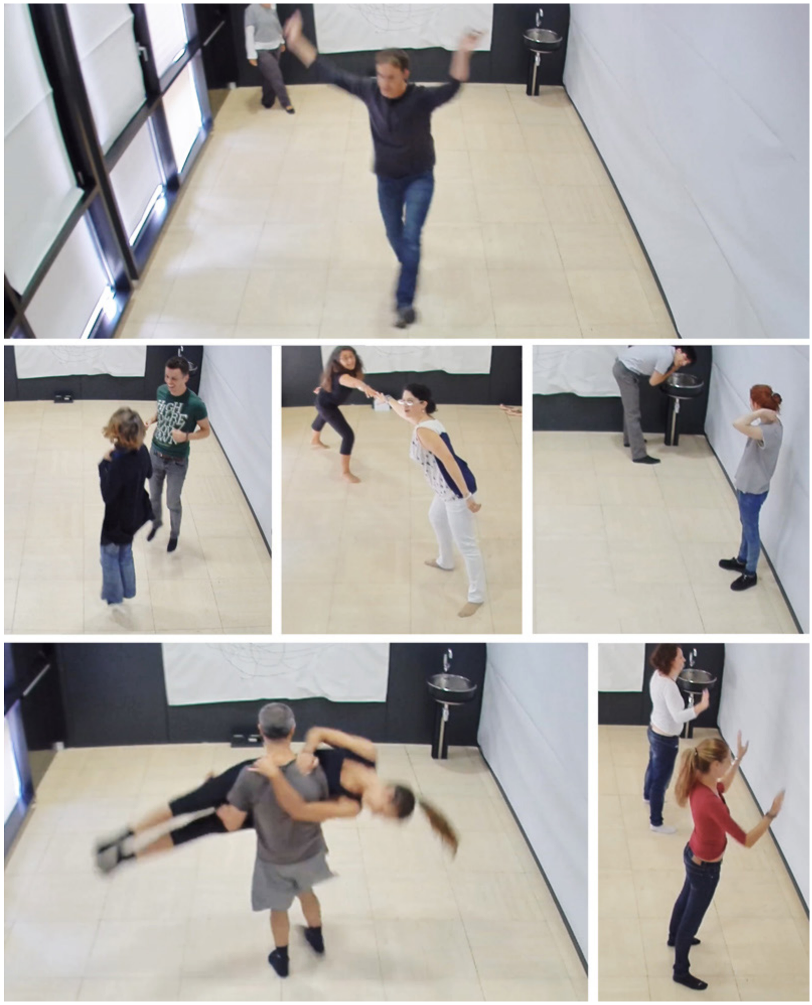


Figure 5.5: Creative interpretations of The Music Room.

Figure 5.5:
The drawing
and the sink
present in
the room.

5.3.5 INTERPRETING NARRATIVES OF USE

The amount of information issued to the visitors determined audience experience. Video analysis revealed that dyads without any information stayed in the room much longer (ranging from 19 to 41 minutes) compared with those who had been given partial and full information (9 to 18 minutes and 7 to 10 minutes, respectively). The uninformed dyads usually spent this extra time in trying to understand the underlying mechanisms of the installation.

Unclear narratives of use proved to foster creative behaviours in the room. Visitors without full information even came up with rules and interpretations that did not actually apply to the installation. "*Making circular movements with hands and arms triggered more harmonic music ... it sounded 'circular'.*" [13p]. As this was not the case, the opposite relationship must have been true: when she perceived the music as "circular", she reacted with circular movements of her hands and arms. Another dyad firmly believed that particular musical patterns were triggered when they were standing in the middle of the room.

Other creative interpretations emerged from the interaction of the visitors with the external objects of the room (Figure 5.5). Three out of four dyads without information spent several minutes interacting with the drawing by touching and waving at it, reckoning it had been placed there for a specific reason. "I think there is some eye-tracker or something behind it...I tried to interact with it" [9i]. One commentator provided a very creative interpretation. He claimed that the drawing was the key to unlocking the system functionality (Figure 5.6). "The drawing was a representation of the whole experience. It gave us a key to interpret the dynamics of the installations. The open box means that the room opens and welcomes us. Then, all those circles and the two hands are there to invite people to move" [2i]. He also repeatedly tried to interact with the sink because he assumed it was somehow connected to the music.

Some commentators discussed how the information they were issued determined their experience. One with partial information said she would have preferred to have received no information because the fun was in the exploration. Four dyads with different levels of information attributed engagement to the lack of rules: "Rules kill creativity" [10i]; "The fact that you do not dictate rules but you just give hints...means that everything is allowed" [4i]. By contrast, the lack of rules negatively impacted on the experience of several visitors: “I felt frustrated. It was like having an instrument and not knowing how to play it” [13i]; “Had you given us information I would have created more” [12p]; “Not knowing the rules does not help people to get into a creative mode” [10p]; “Now that the rules are clearer, I would dare to experiment more with composition.” [10i].

5.3.6 IDIOSYNCRATIC INTERPRETATION

The engagement with The Music Room was as varied as the visitors who enjoyed it. In particular, the relationship between the two members of the dyad and their professional backgrounds strongly influenced their experience.

Those on close terms (i.e. either partners or best friends) visibly enjoyed the presence of their partner: “We enjoyed a lot having to decide together what to do” [10i]. While combining together the words during the concept mapping phase, a commentator reported: “I felt as if I had been in empathy with my partner, because we were playing and dancing together” [6p]. By contrast, the dyad of strangers reported some sort of embarrassment while being close. “Since I have never met her before, I won’t approach her closely, I won’t hug her. I just get close to her a bit but I will always keep my safety margin and she will keep hers” [7i].

Professional background and interests were among the factors that contributed to shaping the experience. Those with a scientific background quite often had a rational approach: “We were keeping a fixed variable while changing the other to understand the mechanism” [13i]. A dyad of physicists felt like “guinea pigs” used for an experiment and indeed, despite being partners, they looked uncomfortable in the room. Conversely, artists generally reported an “introspective experience where I could freely express myself” [13p]. Differences in background and interests deeply influenced visitors’ behaviours even within the same dyad. In two cases, dyads of best friends, both with very different backgrounds, had quite opposite experiences. One was completely focused on understanding the underlying mechanisms and felt frustrated for not being able to do so; her partner, instead, had a very pleasant time dancing and losing themselves in the experience. In another case, one was in a rational mode while the other felt relaxed and immersed in the music.

Apart from personal traits, the variability of these comments may be attributable to the commentators' expectations. One of the dancers tried a wide range of dance movements, because she thought she had been specifically invited by virtue of her dancing skills. One dyad, speculating on the name of the installation, assumed they were about to try to compose music in a recording studio. Lastly, one commentator had expected to have a higher degree of control on the composition so he felt quite disappointed when he realised that fine movements were not taken into consideration.

Additionally, most of the commentators freely provided suggestions about potential applications of The Music Room, and they often visualised possible applications in their domains of interest. They also mentioned several people who could benefit from this system, e.g. children, people with autism, and disabled and visually impaired individuals.

5.3.7 DISCUSSION

This study confirmed that The Music Room consistently meets with visitors' approval. All commentators reported largely positive opinions. This study also offered interesting reflections to understand participants' engagement with The Music Room. In particular, it disclosed that different narratives of use widely affected participants' experience from two perspectives.

1. Ambiguous information proved to be crucial for fostering creative interpretations of the interactive installation. This result confirmed findings of previous works conducted in the broader design area (Höök et al., 2003; Gaver et al., 2004; Dix, 2007). When unbiased by any explanation, visitors created their own interpretation, attributing personal meanings to the system (e.g. a puzzle game, a contemplative space, a dance room, or a playground). They were provided with a *sandbox experience* that was interpretatively flexible (Sengers & Gaver, 2006): after being supplied with a set of tools and mechanics, they were allowed to experiment with the system regardless of the designers' goals. This result suggests that avoiding clear narratives of use has the highest potential for fostering creative engagement: the audience is encouraged to find personal and unexpected meanings and interpretations. On from other hand, such participants' experiences are potentially different to what was originally envisioned by the artist. Also, ambiguous information

has the potential drawback of negatively influencing the experience of some visitors who might be *frustrated* by not precisely knowing how to interact with the system.

2. Different narratives of use also contributed to defining the elements that kept visitors engaged with the installation: *flow* and *exploration*. These two aspects seemed to be mutually exclusive, as they hardly coexisted: visitors were either focused on understanding the functionality of the system or losing themselves in the experience. In general, those who did not receive precise explanations tended to engage more in exploring the system. Most of the time they spent the entire session trying to understand the functionality of the system until they found it, or until they gave up. By contrast, several individuals who received full or partial instructions reported being immersed in the experience. In this case, they did not have to discover the functionality of the system, so they generally spent large parts of their session in *flow condition*.

Besides narratives of use, differences among participants were confirmed as a source of unpredictability of The Music Room. In particular, different combinations of personal traits, background, expectations, and relation between the partners determined the way people responded to The Music Room and how they perceived their experience. Understanding and anticipating this variability is crucial to the interpretation of engagement with an interactive system, and it allows designers to understand the multiplicity of meanings, which often might not match their initial interpretation. In particular, audience expectations strongly affected engagement (Bilda et al., 2008). For instance, if they planned to make music, they would only be able to fulfil their goal when adequately informed about the functionality of the installation. This was the case of the commentator who felt frustrated because the outcome of the experience did not match her expectations and intentions.

The varieties of narrative of use and participants' personal traits contributed to uncovering the multifaceted nature of the installation, which comprises a number of different experiences (Sengers & Gaver, 2004). This characteristic refers to the potential of The Music Room to offer varied experiences to diverse people. Our results suggest that the combination of personal traits and amount of information issued caused this variability to collapse into one particular experience: serving as a collaborative instrument, a support tool for choreographers and therapists, an augmented playground, etc. This understanding opens the way for new design spaces for The Music Room. If we were to offer a single experience to our visitors, this variability should collapse into a single possible experience: this would entail some modifications to the system. For instance, if we set ourselves to create a new tool for choreographers, finer movements would need to be detected and mapped with music and with visual effects, as well. Conversely, if we were to continue pursuing the creation of an intuitive system for music making, we should grant users a higher level of control over the composition. We believe that this finding might prove useful to researchers and artists working in this area.

Finally, some comments on the evaluation methodologies:

Interviews with invited commentators. The interviews were confirmed as an excellent medium through which to gain an in-depth understanding. The interpretation provided by the commentators helped us to better determine how the general public could perceive the installation from the perspective of (i) understanding engagement with an interactive installation, and (ii) generating new design scenarios. Even though hundreds of people had already interacted with the installation during the previous events, many of these reflections had never emerged before.

Concept mapping. The concept mapping method proved particularly useful for helping visitors to stay focused on their role as commentators through a playful task. It showed a good potential for fostering creative interpretations and stimulating reflections. In particular, it enabled commentators to visualise connections by creating *constellations* of meanings and to foster an introspective analysis of their own experience.

Video analysis. Video analysis succeeded in identifying the behaviours of the dyads in the room and their *creative interpretations* of the system. Analysing video footage also proved to be particularly effective to infer the engagement of the visitors by analysing their movements and facial expressions. Finally, confirming findings from the field study, this technique contributed understanding the general experience of the visitors with respect to the envisioned scenarios, confirming that the dancing scenario was most visitors' favourite.

5.4 CONCLUSION

The Music Room was designed for the purpose of providing all users with a novel experience of music making. However, the evaluations that followed the first exhibitions suggested that the audience experience often differed from our expectations. Integrating evidences collected through an array of evaluation techniques disclosed a number of behaviours exhibited in The Music Room, but failed to explain the significance of a number of recurring behaviours exhibited by the visitors. Whether these behaviours were simple expressions of engagement or intentional actions performed to direct the emotion of the music remains to be explored in more depth.

This result prompted us to further inspect this issue by conducting a study with invited commentators. The evaluation of their experience, which was performed with an integration of in-depth interviews, concept mapping and video analysis, proposed that the motivations that predispose participants to perform particular behaviours in The Music Room can be attributable to the amount of information issued to them and to the personal traits of the visitors. In particular, the more information that is given, the closer the experience will be to that originally envisioned by the designer or artist, to the detriment of their participants' creative engagement. Avoiding clear narrative of use has the highest potential for fostering creative engagement, as visitors are more likely to experience personal outcomes and to engage in unforeseen interactions.

Furthermore, with respect to the originally envisioned scenarios (i.e. *dancing*, *acting* and *composing*), these two studies have confirmed that The Music Room ultimately offered a much wider range of non-ordinary and engaging experiences, ranging from music playing to dancing, from enjoyment to relaxation.

A design framework of musical interfaces

THIS CHAPTER PRESENTS MINUET, A FRAMEWORK FOR MUSICAL INTERFACE DESIGN GROUNDED IN THE EXPERIENCE OF THE PLAYER. MINUET AIMS AT PROVIDING NEW PERSPECTIVES ON THE DESIGN OF MUSICAL INTERFACES, REFERRED TO AS A GENERAL TERM THAT COMPRISES DIGITAL MUSICAL INSTRUMENTS AND INTERACTIVE INSTALLATIONS.

6

6.1 INTRODUCTION

For several years, NIME and related communities have been coping with the challenge of framing a design space for musical interfaces. Framing a design space serves the purposes of designers, providing them with concepts that systematise their thinking and stimulate thoughts (Hornecker & Buur, 2006). This practice can inspire design, reveal potential problems (Benford et Al., 2005), and act as a reference point for future studies on user experience (Hornecker & Buur, 2006). This chapter aims at framing a conceptual understanding of musical interface design centred on the experience of the player. The outcome of this work is MINUET (Musical INterfaces for User Experience Tracking), a design framework for the identification of the elements involved in the design of interfaces for musical making. This framework, grounded on related work (Section 2.2.1), is supposed to provide a model for reducing the complexity of the design space of different player experiences, ranging from Disklavier pianists to visitors of The Music Room. MINUET is structured into two stages, following each other in a non-invertible sequence. The first stage analyses the goals of the interface, and the second stage specifies how designers can actually achieve these goals.

MINUET can serve the following purposes: (i) reducing the complexity of the design space of interfaces for music making; (ii) specifying a set of success criteria; and (iii) guiding evaluation procedures. This chapter only focuses on the first point in this paper, leaving the analysis of the other two points to future works. To achieve this goal, we delivered a design process to tackle the issue from the perspectives of designer goal and specifications. The design method is presented in Sections 6.2, and detailed in Section 6.3 and Section 6.4. The reliability of the framework is discussed by drawing parallels with existing research (Section 6.5) and by testing it with the case study of The TwitterRadio (Section 6.6). The final part of the paper provides an insight into future works.

6.2 FRAMEWORK DESIGN

MINUET results from a synthesis of the related work and concepts developed by the author.

6.2.1 DESIGN METHOD

The final structure has been developed over the course of several iterations, including a workshop attended by a selected group of three HCI researchers, a professional musician, and an interactive artist. At first, the related work discussed in 2.2.1 was identified by means of a meta-review process. Related work was primarily collected in the basis of on the references analysed in (Drummond, 2009; Marquez-Borbon, Gurevich, Fyans and Stapleton, 2011), while remaining papers were chosen on the basis of their relevance to the topic. Next, the selected works were schematized in connection to the proposed dimensions (or heuristics) and to the related category of musical interface (DMIs and interactive installation). Later, we organized a one-day workshop: participants were encouraged to discuss affinities and connections among these works. The methodology adopted is affinity analysis (Beyer & Holtzblatt, 1997): under the guidance of the author, participants explored connections, similarities and differences among these works by using sticky notes and several cardboards. They were then asked to propose and discuss a number of dimensions that could correctly cluster the associated design guidelines from the perspective of the experience of the user. From this iteration, three sets of concepts emerged: *Objective of Interaction*, *Constraint* and *Context*, that respectively answered to

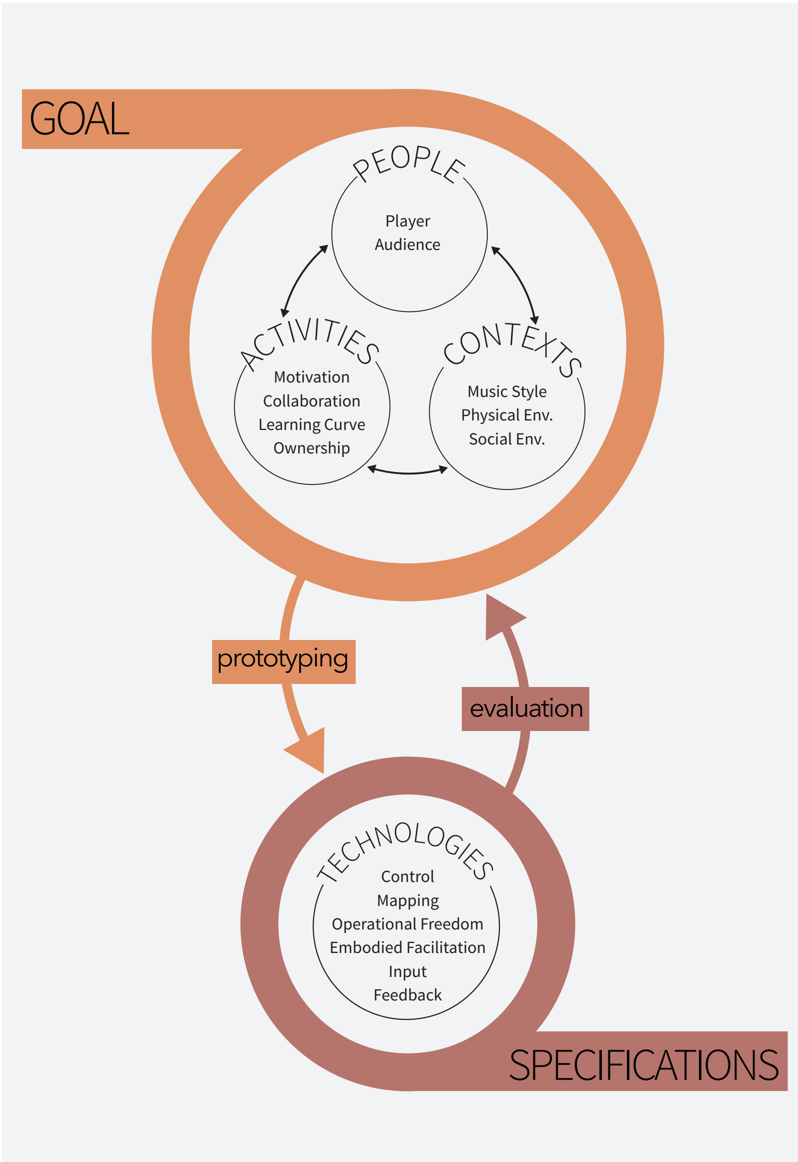


Figure 6.1: MINUET framework.

the questions: *what* the goal of the interface is, *how* this should be designed, *where* and *when* it should take place. The successive version of the model was rearranged by the author and two colleagues, who redistributed the three sets of concepts into two stages of a design process: *goal* and *specifications*. This change was then analysed and supported by some of the workshop participants. Lastly, a few minor modifications to the framework were made during the analysis of the case study reported in Section 6.6.

6.2.2 MODEL

MINUET structures the designing of musical interfaces into a two-stage process: the first stage, *Goal*, describes the objectives of the interface, while the second stage, *Specifications*, help designing the interaction in order to fulfil these objectives (Figure 6.1). Each stage is composed of some entities that address design issues on a more pragmatic level. To model these entities we drew up inspiration from the influential work by (Benyon et Al., 2005), which defines a conceptual model for the design of interactive systems. The PACT framework helps designers to investigate the design process by means of a user-centred technique based on *People*, *Activities*, *Contexts* and *Technologies*. The four entities detail the objectives and the constraints of the interface, and are specified by a series of directly applicable design perspectives (the elements included in the inner circles in Figure 6.1).

The design process starts from the analysis of the designer *goal*, which describes the purposes of interaction. This stage is articulated in three parts: **People,** identifying the end-user (*who*) **Activities,** specifying the kind of interactions the designer has in mind (*what*) and **Contexts** detailing the specifications of the environment (*where* and *when*). Once the interaction goals have been defined, designers can move on to the *Specifications* stage by prototyping the interface. This stage analyses *how* to design the interaction resembling the last entity of PACT framework, **Technologies.** Unlike Benyon's "Technologies", we focus on the identification of interaction requirements rather than considering hardware and software implementations. Lastly, designers can evaluate the proposed specifications by referring back to the original goal.

6.3 GOAL

The first stage of MINUET frames a conceptual model of the interface goal, in the shape of a very high-level user story (Carroll, 2000). The related concepts guide designers to reflect upon the kind of experience they wish to offer. Players can use a musical interface for a number of reasons: to perform, to compose, to control or to learn music. The interface can provide players with a playful experience, stimulate creativity and convey a meditative experience; also, it can have an informative meaning, and it may or may not have educational purposes. In order to cope with the diversity of interactions, this design space is simplified detecting patterns and suggesting insights that can assist designers' reflections. In fact, they are invited to look at their goal through the lenses of People, Activities and Contexts. *People* looks at the designer's objectives from the viewpoint of the targeted category of players and from the role of the audience (*who*). *Activities* questions *what* the envisioned interaction is, by framing the type of musical interface. *Contexts* investigates the environment and the set-up of the interface (*where/when*). The relevance of these entities varies according to the nature and the goals of the interface and the priority scale has to be given by designers themselves.

6.3.1 PEOPLE

This entity specifies the subjects who will engage with the interface, namely players and audience.

Player. A critical point when constructing specific experiences is to define the target user. The interface could address professional or amateur musicians, players that are quite familiar with music, or non-experts. An augmented bow can be designed to enhance the playing experience of professional cello players, or to ease access to the instrument to newcomers.

Audience. In interactive systems the audience may not perceive the interactions between the performer and the system, as they would do with traditional instruments (Gurevich & Cavan Fyans, 2011). Designers that intend to allow spectators to understand such interaction specifically target this issue. The intelligibility or the ambiguousness of the performance greatly depends on the performer's capability of showing others how action-reaction mechanisms work. In some interactive systems, spectators can also actively take part in music creation.

6.3.2 ACTIVITIES

The objective of interaction can be considered from the lens of envisioned player activities. The associated concepts are: motivation, music style, learning curve, ownership and collaboration.

Motivation. This concept provides insights into the motivations of the players, analysing the relevance of the sonic and musical medium with respect to the user experience. In interactive installations, sound often has an accessory goal, as it is supposed to guide the exploration of the environment. Vice versa, the creation or the manipulation of sounds is the primary goal of traditional DMIs and interactive sonification systems, in which users explore particular information conveyed into abstract musical form.

Learning curve. The learning curve provides information on the time required to gain skills with an interface and/or to understand how interaction works. Traditional complex interfaces requiring a significant investment of time should preferably allow high ceiling, as to ensure long-term engagement and enable players to develop skills, as it is the case with traditional instruments. By contrast, museum installations have to be experienced in a very short time, and therefore they must be able to reveal how they work within seconds. In this case the learning curve needs a low floor.

Ownership. Interfaces that specifically address the generation of creative artefacts should aim at the uniqueness of musical expression. When the musical output is a form of self-expression, the execution, the identification and the recognition of unique playing styles should then be supported (Wallis et Al., 2011). By contrast, when the interaction is not centred on the actual production of a musical output, replicable outcomes are to be expected. Ownership can also refer to the way players interact with the interface if the system allows them to configure the interface according to their own wishes.

Collaboration. Musical activities can be carried out alone or in a social context. Designing collaborative experiences means considering awareness of others, and factors such as synchronization and coordination. Sometimes, musical interfaces can only be fully exploited in collaborative contexts: in these cases, specific efforts need to be made to ensure the expected social dynamics.

6.3.3 CONTEXTS

Contexts consider the environment in which the interaction takes place, i.e. all those elements that can help the identification of the interaction goals. This entity may not be very important for traditional DMIs, yet it is particularly relevant for installations and collaborative interactive systems. The associated concepts (music style, physical environment and social environment), detail what makes an application unique in relation to similar works.

Music style. In novel musical interfaces sound is produced by the computer, so the associated musical style is not influenced by the acoustic features of the instrument as in traditional music. This feature opens up new horizons for designers, who can choose the music style that better suits their goals, irrespective of the nature of the device. Identifying a musical style is crucial when it comes to targeting a specific population. For instance, when confronted with a broad target population, designers should make use of a familiar music language, such as tonal music. On the other hand, when targeting a community of experts or avant-garde musicians, the music itself can be experimental. Addressing a specific musical style genre can also be the primary objective of the interface (e.g. the musical controllers for electronic music such as Korg Kaossilator 2)[1].

Physical environment. When reflecting on the goals of the interface, its own nature frames the importance of the space in which it is going to take place. While traditional controllers do not require a precise spatial configuration, it is important to consider the physical environment hosting the installation. Some installations are designed to be exhibited on the occasion of specific events while others are hosted by museums for long periods.

Social environment. The social context and the proximity between players mirror different kinds of interpersonal relations. Intimate experiences need a small, secluded space, while provocative installations can either be open to the greatest number of passers-by or force individuals to interact with an object on a one-on-one basis. Designers might want to draw the attention of museum visitors, or to arouse the interest of people and of the media into a specific social issue.

1 *http://www.korg.com/us/products/dj/kaossilator2/*

6.4 SPECIFICATIONS

The second stage of the design process collects the goal considered by the previous issue and reflects on the interaction constraints between the player(s) and the system. When designing for intuitive experiences the amount of interaction possibilities should be restricted, as to guarantee an easy access to the players. By contrast, musical controllers should enable players to manage a multitude of parameters, as to have full control on the generated music. Specifications must be considered according to the level of control and to the input and feedback modalities as well as to the presence of human facilitators that might result in a loss of operational freedom.

The degree to which the player controls the music varies along a continuum, which includes the different categories of interfaces. For the sake of convenience, though, we will discretize three categories: low, medium and high control. Low-level controllers provide control on each note or on specific sonic parameters, just like traditional instruments. They usually address an expert population, as they rely on musical notations that are unknown to non-experts. Mid-level controllers give access to higher musical structures at bar-level or score-level such as rhythmic pattern, melody direction, mode, musical processes and loops. Depending on their closeness to note-level, they might be more or less accessible to non–experts. High-level controllers operate outside the musical domain, so they are generally open to everybody but do not allow any subtle interaction with musical structures.

Mapping. The level of control influences the complexity of mapping strategies. Low-level controllers require a convergent mapping (Hunt & Kirk, 2000): in order to produce a single pitch, a sequence of physical tasks needs to be performed just as playing a note on the guitar usually requires the synchronization of two or more fingers). Vice versa, in interfaces controlling high-level musical elements, a single *mediated* parameter affects many musical factors - divergent mapping (Hunt & Kirk, 2000). For instance, controlling *mode* influences two lower-level musical parameter, i.e. harmonic and melodic intervals. Sometimes interface metaphors are needed to hide mask music complexity from the player's interaction space. Interface metaphors indeed "allow users to readily make inferences about how to operate unfamiliar user interfaces by mapping existing skills and knowledge from some familiar source domain"

(Wilkie, Holland and Mulholland, 2013). In this case, a divergent mapping is necessary: controlling the emotional character of a piece will influence several musical parameters (e.g. mode, tempo, melodic direction and volume).

Input. Players can interact with the musical interface through symbolic, para-linguistic, involuntary and subconscious modalities (Bongers & van der Veer, 2007). The production of sounds might require physical energy an ergotic gesture. Alternatively, players could interact on a semantic level, through visual, tactile or semiotic gestures.

Feedback. This concept focuses on the presence and the role of feedback modalities (auditory, visual, tactile, or kinesthetic – Birnbaum et Al., 2005). Designers are invited to consider the effects related to different feedback modalities. Adding multi-sensorial feedback can augment the experience, but also distract a performer. In most cases, the visual feedback is an accessory factor but sometimes it can become the main focus of user experience (Fels & Mase, 1999).

Operational freedom. Operational freedom defines the potential of players to express a creative interaction with the system and the flexibility of the interaction. Designers seeking to stimulate creativity, improvisation and adaptation should envision flexible and constraint-free interactions, while educational or training-oriented interfaces are better achieved through rigid task-achievements procedures. A snare drum has rather limited interaction possibilities, but offers endless compositional options. Vice versa, The Music Room does permit players to creatively interact with the system, but the musical control depends on the choice of the piece's emotional tone.

Embodied facilitation. Embodied facilitation refers to the physical configuration of the interface, as well as to the presence of human facilitators. The design of the interface can impose limitations if it "facilitates, prohibits and hinders some actions, allowing, directing, and limiting behaviour" (Hornecker & Buur, 2006). In this case, the designer suggests a particular strategy or interaction trajectory using constraints. Interaction constraints could also be forced by human and virtual facilitators that pilot the interaction. Facilitators can show players how to interact with the interface and other players and how to accomplish tasks. If needed, the spatial environment should support the envisioned interaction. When designing collaborative installations, the physical set-up should encourage players to collaborate by constraining or facilitating their interaction, while a round space where players have the same distance from the object of the interaction would be appropriate for shared-control interfaces.

6.5 RELIABILITY OF THE FRAMEWORK

In order to argument the proposed framework, in this section we show how previous frameworks and taxonomies are accommodated within it. In Table 6.1 the main dimensions of the related work are associated with the elements proposed by MINUET.

Table 6.1: MINUET compared to related work (the first letter refers to the associated PACT entity).

Related work	Dimension	MINUET
Birnbaum et Al., 2005	Required expertise	P: Player
		A: Learning Curve
	Musical control	T: Control
	Feedback modalities	T: Feedback
	Degrees of freedom	T: Operational freedom
	Inter-actors	A: Collaboration
	Distribution in space	C: Physical environment
	Role of sound	A: Motivation
Wallis et Al., 2013	Incrementality	A: Learning Curve
	Complexity	
	Immediacy	
	Ownership	A: Ownership
	Operational Freedom	T: Operational Freedom
	Demonstrability	P: Audience
		P: Player
	Cooperation	A: Collaboration

Blaine & Fels, 2000	Physical device	T: Input
	Player interaction	
	Learning curve	A: Learning curve
	Pathway to expertise	A: Ownership
	Physicality between players	A: Collaboration
		T: Embodied facilitation
	Musical range	T: Control
	Media	T: Feedback
		A: Motivation
Overholt, 2009	Gestures intuitiveness and perceptibility	P: Player
		T: Input
	Mapping richness	T: Mapping
	Range of expression	T: Control
Hunt & Kirk, 2000	Mapping strategies	T: Mapping
Cumhur Erkut & Disçioglu., 2011	Modality of interaction	T: Input
Johnston et Al., 2008	Modes of interaction	T: Input
Wanderley & Orio, 20002	Learnability	A: Learning curve
	Explorability	T: Operational Freedom
	Feature controllability	T: Control
	Timing controllability	
Jordà, 2004	Balance	A: Learning Curve
	Instrument efficiency	T: Control
		T: Operational Freedom
	Playability	A: Learning curve
	Learning curve	

6.6 TWITTERRADIO, A CASE STUDY

This section tests with a case study the validity of MINUET to guide researchers throughout the design process. To this end, we introduce the conceptual design of The TwitterRadio, an interactive installation designed to convey public opinions on globally trending topics in a suitable musical form.

Every minute, thousands of multimedia elements are generated, for the purpose of sharing life experiences and feelings with friends, and voicing personal opinions on trending topics. These data are usually explored through a visual approach, involving the use of pictures, videos or graphs. The main idea behind The TwitterRadio is to express these data with music, given its high potential for expressing moods. Specifically, the visitors of The TwitterRadio would browse a list of trending news and listen to the mood of public opinions on those specific topics. The adopted data source is Twitter, which numbers more than 280 million active users (by the end of 2014) who constantly share their thoughts and feelings on personal and social issues. The idea is to collect all recent tweets labelled with trending hashtags and to retrieve information about their emotional valence and popularity. These features can be mapped to music in order to create melodies that match the mood of the tweets. A rarely mentioned hashtag, for instance, would result in a slow, flat melody, while a trending hashtag would result in intense sounds. MINUET was adopted to elaborate this idea, developing a comprehensive conceptual design.

6.6.1 GOAL

The original objective of The TwitterRadio was elaborated using the concepts proposed by MINUET in the *Goal* phase.

Activities and People. The TwitterRadio provides a novel environment for experiencing tweet-based contents in an auditory form. The complexity of the interaction needed to be reduced to a minimum, given the walk-up-and-play character of the installation (*learning curve*), intended to be experienced in museums or art galleries (*social environment*). The TwitterRadio indeed targets the same audience of The Music Room: a general public that potentially lacks musical training (*player*). The originally envisaged interaction was limited to selecting existing themes. To increase the active engagement of visitors, we decided to allow them to explore themes they are interested in by typing in theme-related hashtags. The *Motivation* concept also prompted the consideration of other facets of the installation to conceptualise further interaction scenarios. A number of scenarios were envisioned during a design workshop with 12 user experience researchers. The outcome of this design process resulted in the proposal of a new scenario: the idea was to exploit the social character of The TwitterRadio to foster critical thinking in the listeners. We were expecting the emotional character of the music not to match listeners' expectations at times. In these cases, they could be actively involved in discussions with other listeners *(collaboration)* to understand, for instance, *why does Christmas sound so sad?*

Contexts. The physical design and set-up is particularly important for this installation. First, the *social environment* in which it is located is crucial. The idea was to base The TwitterRadio on an analogy to traditional radios, from an aesthetic and a functional perspective. People used to gather in front of the radio to listen to and comment on news from the world or to dance to the music played by radio stations. The TwitterRadio combines both modalities into a single experience. However, we mainly intend The TwitterRadio as a news feed player rather than as a music player. To this end, it has to be experienced by a limited number of visitors at once; thus, it needs to be set up in a quiet, meditative location. The *physical environment* has to guarantee the visibility of the installation from a close point of view. Concerning the *music style,* given that The TwitterRadio targets the general public, the music has to be in a style everyone is familiar with. As a consequence, similar to The Music Room, the music was selected to be of a classical-like style.

6.6.2 SPECIFICATIONS

Once the goals had been outlined, the second stage focused on prototyping possible design implementations. To interact with the music, visitors will operate though a semantic, high-level *control* of the music: they can only influence it by selecting a topic of interest. A high-level control requires the adoption of an algorithmic composer that transforms collected semantic information into combinations of musical parameters (convergent *mapping*). To this end, Robin was utilised to generate the music played by The

TwitterRadio. Sound is then the main *feedback* of the installation. Besides the actual music played by the radio when tuned to specific stations, to improve the analogy to old radios Robin generates radio-reception-noise during transition states. Furthermore, a visual feedback needs to be implemented to display the list of the stations and to indicate the current station.

The user interaction with the system is limited to selecting pre-existing topics, or to typing in new ones. Following the design objectives of resembling an old radio, visitors will physically interact with the system through a knob that plainly reveals the interaction strategy (*embodied manipulation*). The concept of *operational freedom* opened up reflections about possible creative interactions with the installation. In addition to this interaction, visitors could subvert the original idea, appropriating the installation as a musical controller. Visitors could indeed operate on the knob in a creative manner instead of simply using it for selecting their favourite theme. For instance, they could (i) rotate the knob in and out of a theme to rhythmically alternate noise and music, (ii) try to create a *song structure* by purposely switching between themes with different moods, or (iii) select a theme with a mood that is aligned with theirs.

6.6.3 Discussion

MINUET contributed to outlining the design objectives and specifications of The TwitterRadio from a number of perspectives. First, the *goal* stage helped to frame the character of the installation. Initially, it was intended as an exploration of trending news in an auditory form. Then, elaborating on the *motivation* concept, we further expanded the nature of the system to eventually include the *critical design* aspect of it. Furthermore, this stage helped to identify design requirements relative to the look and the set-up of the installation. Second, the *specification* stage contributed to elaborating possible interaction trajectories with the systems and detailing their implementation. Detailing design implementations opened new interaction possibilities that were not envisaged at the initial design stage. In particular, the presence of the knob and tuning noise suggested possible creative exploitations of the system.

6.7 CONCLUSION

A general consensus suggests that a technology-oriented design well suits musical controllers that resemble the physicality and the objectives of acoustic instruments (Miranda & Wanderley, 2006). If we considered musical interfaces as an umbrella term that also includes interactive installations (Wanderley & Orio, 2002; Birnbaum et Al., 2006) however, this approach may prove inconsistent, due to the diversity and the complexity of this design space. Designers working on this category of interfaces, indeed, tend to focus on a particular user experience rather than a precise set of musical activities. Performing formal design and evaluation strategy might prove out to be very demanding.

Through a systematic review and a reinterpretation of related design dimensions and heuristics, this chapter proposed a new conceptual model that aims at tackling this issue. The fruit of this research, MINUET, is meant to stimulate reflection on both an abstract and a practical level when designing novel musical interfaces. As well as incorporating all of the varieties of musical interfaces (addressing this design space and leaving behind the somewhat artificial boundaries intrinsic to any definition), MINUET offers two important contributions. First, it shifts the focus of previous frameworks, instead framing a design space of musical interfaces centred on players' experiences. Second, to the best of our knowledge, this is the first attempt to develop a framework with a precise temporal unfolding rather than a set of heuristics.

This study laid the foundations for developing a framework for the design and evaluation of musical interfaces. So far, we have provided a conceptual model that frames the designer's goals into a set of specifications. The next step will be to investigate how these specifications can be validated through a series of success criteria. Subsequently, we will conduct a study to examine how to gather empirical data on the basis of success criteria in their natural setting. Indeed, due to the hedonic nature of the interfaces and due to the fact that they are often used in non-controlled live conditions, most of the traditional HCI techniques for evaluating users experiences cannot be applied to this design space.

Finale

THIS CHAPTER SUMMARISES AND DISCUSSES THE RESULTS, CONTRIBUTIONS AND IMPLICATIONS OF THE THESIS; POINTS TO LIMITATIONS; AND PROPOSES FUTURE WORK.

7

7.1 THESIS CONTRIBUTIONS

This thesis contributed to three research areas: musical interface design, psychology of music, and algorithmic composition. In this section, we summarise the individual contributions.

7.1.1 MUSICAL INTERFACE DESIGN

This thesis proposed advancements in the development of interfaces for music making that target musically untrained users. With respect to existing categorisation of musical interfaces (Wanderley & Orio, 2002; Birnbaum et al., 2006), this category roughly lies in between DMIs and interactive installations. A limited number of DMIs have been trying to ease music making since the last decade (Wanderley & Orio, 2002; Miranda & Wanderley, 2006). Their contribution often focused on proposing new interactive solutions to reduce the reliance on specialised motor skills that are necessary to make music with traditional instruments (Blaine & Perkis, 2000; Patten et al., 2002; Jordà et al., 2006). However, in most of the cases these solutions still relied on a deep knowledge of musical notation, which is likely to be meaningless to musically untrained users. That is to say, current proposals offered only a partial solution to the problem.

To address this problem, this thesis proposed to exploit the language of emotion to communicate musical meanings. In other words, musical meanings are encoded into emotions that, in turn, are mapped into combinations of structural factors that determine the affective character of the music. The music control is shared between the computer and the human. Specifically, the human controls the emotional character of music.

Existing interactive installations have been trying to empower the audience to control the music content interactively (Rowe, 1993; Machover, 1996; Camurri et al., 2010; van Boerdonk, 2009, Fabiani et al., 2011). These installations allow visitors to influence the composition by dancing in a stage (Camurri et al., 2003), interacting through tangible objects (van Boerdonk, 2009) or performing gestures with mobile phones (Varni et al., 2010; Fabiani et al., 2010). However, most of these systems do not provide a detailed account of their design process and little is known of prototyping and evaluation stages. As a consequence, their contribution to the design field is limited.

Furthermore, to the best of our knowledge, in this context, nobody ever integrated algorithmic affective composers, which could allow visitors to meaningfully interact with the music. Rather, the user interaction was limited to high-level control modalities, such as editing pre-composed pieces (Camurri et al., 2010) or deciding the level of orchestration (Varni et al., 2010). In some cases users are let to control sonic features (Šimbelis et al., 2014), which are unlikely to convey appropriate meaning to musically untrained visitors. Also, in almost all the cases, the evaluation of the visitors' experience of the players is almost missing.

With respect to them, this thesis offered three main contributions.

1. *Exploiting the metaphor of emotion to make music*. In interactive systems for music making new interactive metaphors need to be developed to mediate abstract musical meanings in concrete domains that everybody can easily understand. This thesis proposed emotion to encode this mediation. Replacing musical notations with emotions in the process of music making results in a number of critical implications in the design process. First, in order to be communicated to the system, emotions need to be encoded into specific media. Second, the involvement of the player changes, as the traditional note-to-note control is replaced by control strategies based on emotions. Lastly, an algorithmic composer needs to be developed to allow users to control the emotional character of an original music in real-time. This system was tested with the case study of The Music Room (Chapter 4). This is a considerable step forward in the vision of empowering the audience of interactive installations to have meaningful control on the musical outcome.

2. *Systematic evaluation of visitor's experience in interactive installation*. A systematic evaluation of the experience of the visitors of The Music Room is provided with a field and a controlled evaluation. The study provided solutions to the issue of collecting and evaluating the experience of the audience in interactive installations (Edmonds, 2010). The result of this study, discussed in Chapter 5, proved that conducting a formal evaluation of an interactive installation could in fact open numerous insights on the experience of the audience and on the status of the system itself. The evaluation methodologies are partially borrowed from broader design areas, such as interactive art (Edmonds, 2010; Bilda et al., 2008; Costello, Muller, Amitani and Edmonds, 2005) and product design (Gaver 2007). Lastly, this thesis contributed to an operational definition of the concept of *engagement* with interactive art.

3. *Developing a design framework of musical interfaces*. MINUET, a design framework of interfaces for music making is proposed (Chapter 6). MINUET has three main advantages with respect to related

work: (i) it overcomes the dualism between DMIs and interactive installations, (ii) it reduces the complexity of the design space emphasizing the experience of the player, and (iii) it provides designers with new perspectives and reflections when designing novel interfaces for music making.

7.1.2 Algorithmic composition

A second research area that might benefit from the findings discussed in this thesis is that of algorithmic affective composition, a recent research area that, to date, counts only a handful of studies (Legaspi et al., 2007; Hoeberechts & Shantz, 2009; Oliveira & Cardoso, 2010; Livingstone et al., 2010; Wallis et al., 2011). These studies contributed defining a novel branch of algorithmic composition, allowing users to influence the emotional configuration of the composition in real time. However, given the infancy of this research area, a number of severe limitations still reduce the actual applicability of these systems. First, the quality of the music generated by all these systems is variable, and seems to be the result of a computational exercise rather than the result of a creative act. Second, a validation study to test the effective capability of the system to communicate correct emotional response in the listener is missing. Lastly, none of these algorithms have been used in actual interactive installations as, so far, users can only interact with these algorithms with clickable interfaces and select their desired emotions by interacting though a mouse.

Chapter 3 described the implementation of Robin, an algorithmic affective composer, specifically developed to overcome the limitations of the related work.

1. *Focusing on the quality of the music.* The design of Robin focused on generating quality music rather than proposing innovative solutions to algorithmic composition. As a result, in the author's opinion, the quality of the music seems of higher standards if compared to the related work. Music liking is a highly subjective feeling, so this claim would need to be supported by formal comparisons with existing systems. However, considering the difficulty of retrieving the code of such systems, this comparative study is left to future work. However, initial evidence indicates that people liked the music generated by Robin. Indeed, the visitors of the Music Room rated the quality of the music with an average of 3.98 on a scale from 1 (*I did not like the music at all*) to 5 (*I liked the music a lot*). This result is

particularly positive if compared to the ratings given to the music excerpts composed by a human, which were used in the experimental study described in Section 3.2. In that case, participants rated the *liking* of the excerpts with 3.66 on a scale 1-7, which corresponds to 2.61 on a scale 1-5.

2. *Validity evaluation.* The actual capability of Robin of communicating correct emotions in the listeners was tested in an experimental study. Participants were asked to rate the perceived valence and arousal of 20 snippets generated by Robin, divided in four combinations of valence and arousal. Results indicate that Robin can accurately communicate arousal levels in the listeners. Valence perception resulted more complex and tend to improve when different structural factors (e.g. tempo and mode) converge. This finding can be explained in the lights of the difficulty of non-musicians to discriminate divergent stimulation (see result of the experiment at Section 3.2).

3. *Exploiting Robin in interactive installation.* Robin is the first attempt to exploit an algorithmic affective composer in context of interactive installations. The effectiveness of Robin to be used to generate music in this context was tested in two interactive installations, The Music Room and The TwitterRadio.

7.1.3 Psychology of music

This thesis also offered some important contributions to the psychology of music. The study presented in Section 3.2 focussed on understanding the role of expertise when evaluating the influence of tempo and mode on emotional perception. Existing studies proposed that the perception of emotions in music does not vary between musicians and non-musicians (Robazza et al., 1994; Bigand et al., 2005). However, to date, no studies have been trying to systematically analyse the effect of tempo and mode on the perception of valence and arousal taking into consideration listener's expertise. Furthermore, related works do not agree on the emotional perception in case of divergent conditions of mode and tempo (Section 2.3.4). The work presented in Section 3.2 aimed at filling these gaps with an experimental study aiming at understanding whether and how listener's expertise influences the perceived valence and arousal to tempo and mode alterations.

Furthermore, to date, there is not a single correct alternative when selecting stimuli for experimental studies on music and emotion (Gabrielsson & Lindström, 2010; Bigand et al., 2005; Robazza et al., 1994). In some cases, stimuli are selected among existing pieces of music. In other cases, stimuli are composed ad hoc as short sound sequences where individual factors are systematically manipulated. The former approach ensures an ecological validity of the music but conclusions regarding the effects of individual musical parameters can be only tentative (Gabrielsson & Lindström, 2010). By contrast, the latter approach guarantees a systematic analysis of the effects of the probed factors, but the ecological validity of the music is reduced (Gabrielsson & Lindström, 2010). Also, with this approach, the similarities between each other stimuli could disengage participants from the listening activity thus reducing the validity of the results.

With respect to these limitations, this thesis offered three main contributions.

1. *Probing into the role of expertise in perceiving emotions in music.* Expertise has a role in influencing listener's judgment of emotion in music, but only for the dimension of valence and in case of divergent stimulations. Non-musicians were less sensitive than musicians to the information conveyed by mode and they mainly relied on the emotional information conveyed by tempo.

2. *Understanding emotion perception when tempo and mode diverge.* When evaluating the valence of a piece in divergent conditions, trained musicians, who are primarily influenced by mode, judged the major-slow condition as more positive than minor-fast condition, whereas non-musicians rated both conditions halfway.

3. *Adopting algorithmically composed stimuli for experimental studies.* The experimental studies presented in Section 3.4, besides assessing the validity of Robin of communicating correct emotional response in the listener, suggests its potential of being adopted as *stimuli generator* in this kind of experiments. Indeed, the results of the study correctly replicate previous works findings, and the average liking of the snippets is markedly higher if compared to those created by a human composer for the first experiment (Section 3.2). These findings suggest that the music composed by Robin has an ecological validity and allows a systematic manipulation of the structural factors of interest.

7.2 CONCLUSION AND FUTURE WORKS

This thesis presented a design solution to offer new experiences of music making to an untrained audience. The fundamental point of this solution is the employment of a new interaction metaphor that mediates abstract musical meanings in a concrete domain. The metaphor proposed by this work is that of emotion, given its universal availably and natural connection with music. This design solution was deeply tested with the case study of The Music Room. Results showed particularly efficient if confronted with the engagement of the visitors and the boost of musical creativity they experienced. A question that remains open for future work is whether and how other semantic descriptors could act as mediators of music complexity. Semantic descriptors have been already successfully adopted in music retrieval (Leman, 2008): when selecting songs to add to our play queue, we often make conscious or subconscious queries based on semantic or symbolic descriptors (e.g. memories, emotions, genres and narratives) that "involve a description of the intrinsic qualities of the music as well as of subjective experiences of the music" (Leman, 2008).

This thesis proposed some key ideas that could be considered and applied to the creation of other interactive installations, as well as strategies for evaluating and understanding the nature of such works. Indeed, as the number of artists, designers, researchers, and practitioners engaged in interactive art increases, there is an ever growing need and interest in learning principles for designing effective experiences, understanding how to evaluate visitors' response, and measuring the success of the artwork. The systematic approach taken in evaluating The Music Room expanded formal knowledge on principles of design in interactive art. Results have demonstrated how different amounts of explanation affected visitors' engagement with the installation, opening the door to subsequent research. The considerations offered by this thesis could be indeed applied to other interactive art forms. Furthermore, a number of techniques for evaluating visitors experience with interactive art were reviewed, focusing on benefits and drawbacks.

Robin, the algorithmic affective composer presented in this thesis put forward several contributions in the field of automatic composition of affective music. In particular, the author's opinion is that the quality of the music generated by Robin constitutes a progress with respect to that of related studies (Hoeberechts & Shantz, 2009; Livingstone et Al, 2010; Oliveira & Cardoso, 2010; Wallis et al., 2011). Formal comparative studies are left to future works. Furthermore, the algorithm driving Robin can be further improved. For example, the quality of the music could benefit from (i) enhancing the chord progression matrix by increasing the order of the Markov chain; (ii) developing more complex accompaniment lines; and (iii) developing algorithmic counterpoints. To improve the accuracy in communicating emotions, performative factors such as *articulation, dynamics*, and *timing* could be also included in Robin. This advancement can be gained by integrating Robin with an algorithmic expressive sequencer such as the pDM (Friberg, Bresin and Sundberg, 2006).

This thesis also contributed to the psychology of music by advancing current understanding of the role of expertise on the emotional perception of music. The effect of mode and tempo was systematically tested in the emotional dimensions of valence and arousal taking into consideration listeners' expertise. The results revealed that, as listeners go through formal musical training they tend to rely on mode rather than tempo when evaluating the valence of a musical composition. In case of diverging conditions of mode and tempo, non-musicians were not able to provide a clear indication of the perceived emotions. This result suggests that future research should be conducted to investigate the emotional perception in case of diverging conditions of mode and tempo. Furthermore, future studies testing the emotional perception of structural factors could exploit Robin as stimuli generator: ecologically valid and musically natural music would be combined with systematic control over the tested factors.

Bibliography

Altman, D. G. (1990). *Practical statistics for medical research*. CRC Press.

Amabile, T. M. (1996). *Creativity in context*. Westview Press.

Bahn, C., & Trueman, D. (2001, April). Interface: electronic chamber ensemble. In *Proceedings of the 2001 conference on New interfaces for musical expression* (pp. 1-5). National University of Singapore.

Benford, S., Schnädelbach, H., Koleva, B., Anastasi, R., Greenhalgh, C., Rodden, T., ... & Steed, A. (2005). Expected, sensed, and desired: A framework for designing sensing-based interaction. *ACM Transactions on Computer-Human Interaction (TOCHI)*, *12*(1), 3-30.

Benyon, D., Turner, P., & Turner, S. (2005). *Designing interactive systems: People, activities, contexts, technologies*. Pearson Education.

Beyer, H., & Holtzblatt, K. (1997). *Contextual design: defining customer-centered systems*. Elsevier.

Bharucha, J. J., & Stoeckig, K. (1987). Priming of chords: Spreading activation or overlapping frequency spectra?. *Perception & Psychophysics*, *41*(6), 519-524.

Bigand, E., Vieillard, S., Madurell, F., Marozeau, J., & Dacquet, A. (2005). Multidimensional scaling of emotional responses to music: The effect of musical expertise and of the duration of the excerpts. *Cognition & Emotion*, *19*(8), 1113-1139.

Bigand, E., & Poulin-Charronnat, B. (2006). Are we "experienced listeners"? A review of the musical capacities that do not depend on formal musical training. *Cognition*, *100*(1), 100-130.

Bilda, Z., Edmonds, E., & Candy, L. (2008). Designing for creative engagement. *Design Studies*, *29*(6), 525-540.

Birnbaum, D., Fiebrink, R., Malloch, J., & Wanderley, M. M. (2005). Towards a dimension space for musical devices. In *Proceedings of the 2005 conference on New interfaces for musical expression* (pp. 192-195). National University of Singapore.

Blaine, T., & Perkis, T. (2000). The Jam-O-Drum interactive music system: a study in interaction design. In *Proceedings of the 3rd conference on Designing interactive systems: processes, practices, methods, and techniques*(pp. 165-173). ACM.

Blaine, T., & Fels, S. (2003). Collaborative musical experiences for novices.*Journal of New Music Research*, *32*(4), 411-428.

Boenn, G., Brain, M., & De Vos, M. (2008). Automatic composition of melodic and harmonic music by answer set programming. In *Logic Programming* (pp. 160-174). Springer Berlin Heidelberg.

Boenn, G., Brain, M., De Vos, M., & Ffitch, J. (2011). Automatic music composition using answer set programming. *Theory and practice of logic programming*, *11*(2-3), 397-427.

Bongers, B., & van der Veer, G. C. (2007). Towards a Multimodal Interaction Space: categorisation and applications. *Personal and Ubiquitous Computing*, *11*(8), 609-619.

Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of behavior therapy and experimental psychiatry*, *25*(1), 49-59.

Bradski, G. R. (1998, October). Real time face and object tracking as a component of a perceptual user interface. In *Applications of Computer Vision, 1998. WACV'98. Proceedings., Fourth IEEE Workshop on* (pp. 214-219). IEEE.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative research in psychology*, *3*(2), 77-101.

Brauneis, R. (2014). Musical Work Copyright for the Era of Digital Sound Technology: Looking Beyond Composition and Performance. *GWU Legal Studies Research Paper*.

Bresin, R., & Friberg, A. (2000). Emotional coloring of computer-controlled music performances. *Computer Music Journal*, *24*(4), 44-63.

Bresin, R., & Friberg, A. (2011). Emotion rendering in music: range and characteristic values of seven musical variables. *Cortex*, *47*(9), 1068-1081.

Brooks, F. P., Hopkins, A. L., Neumann, P. G., & Wright, W. V. (1957). An experiment in musical composition. *Electronic Computers, IRE Transactions on*, (3), 175-182.

Burraston, D., & Edmonds, E. (2005). Cellular automata in generative electronic music and sonic art: a historical and technical review. *Digital Creativity*, *16*(3), 165-185.

Camurri, A., Hashimoto, S., Ricchetti, M., Ricci, A., Suzuki, K., Trocca, R., & Volpe, G. (2000). Eyesweb: Toward gesture and affect recognition in interactive dance and music systems. *Computer Music Journal*, *24*(1), 57-69.

Camurri, A., Lagerlöf, I., & Volpe, G. (2003). Recognizing emotion from dance movement: comparison of spectator recognition and automated techniques. *International journal of human-computer studies*, *59*(1), 213-225.

Camurri, A., Volpe, G., Poli, G. D., & Leman, M. (2005). Communicating expressiveness and affect in multimodal interactive systems. *Ieee Multimedia*,*12*(1), 43-53.

Camurri, A., Canepa, C., & Volpe, G. (2007). Active listening to a virtual orchestra through an expressive gestural interface: The Orchestra Explorer. In*Proceedings of the 7th international conference on New interfaces for musical expression* (pp. 56-61). ACM.

Camurri, A., Varni, G., & Volpe, G. (2010). Towards analysis of expressive gesture in groups of users: computational models of expressive social interaction. In *Gesture in Embodied Communication and Human-Computer Interaction* (pp. 122-133). Springer Berlin Heidelberg.

Canazza, S., De Poli, G., & Rodà, A. (2014), "CaRo 2.0. An interactive system for expressive music rendering", Advances in Human-Computer Interaction, (submitted).

Candy, L., Amitani, S., & Bilda, Z. (2006). Practice-led strategies for interactive art research. *CoDesign*, *2*(4), 209-223.

Carroll, J. M. (2000). *Making use: scenario-based design of human-computer interactions*. MIT press.

Castro, S. L., & Lima, C. F. (2014). Age and Musical Expertise Influence Emotion Recognition in Music. *Music Perception: An Interdisciplinary Journal*, *32*(2), 125-142.

Chai, W., & Vercoe, B. (2001). Folk music classification using hidden Markov models. In *Proceedings of International Conference on Artificial Intelligence* (Vol. 6, No. 6.4).

Cooke, D. (1959). *The language of music.* Oxford University Press.

Cope, D. (2005). *Computer models of musical creativity* (p. xi462). Cambridge: MIT Press.

Costello, B., Muller, L., Amitani, S., & Edmonds, E. (2005). Understanding the experience of interactive art: Iamascope in Beta_space. In*Proceedings of the second Australasian conference on Interactive entertainment* (pp. 49-56). Creativity & Cognition Studios Press.

Csikszentmihalyi, M. (1991). *Flow: The psychology of optimal experience* (Vol. 41). New York: HarperPerennial.

Darwin, C. (1859). On the Origin of Species. London: John Murray. Fang,

Dix, A. (2007). Designing for appropriation. In *Proceedings of the 21st British HCI Group Annual Conference on People and Computers: HCI... but not as we know it-Volume 2* (pp. 27-30). British Computer Society.

Downey, J. E. (1897). A musical experiment. *The American Journal of Psychology*, *9*(1), 63-69.

Drummond, J. (2009). Understanding interactive systems. *Organised Sound*,*14*(02), 124-133.

Edmonds, E., Muller, L., & Connell, M. (2006). On creative engagement. *Visual Communication*, *5*(3), 307-322.

Edmonds, E. (2010). The art of interaction. *Digital Creativity*, *21*(4), 257-264.

Eerola, T., Friberg, A., & Bresin, R. (2013). Emotional expression in music: contribution, linearity, and additivity of primary musical cues. Frontiers in Psychology, 4(487), 1-12.

Ekman, P. (1992). An argument for basic emotions. *Cognition & Emotion*, *6*(3-4), 169-200.

Einstein, A. (1905). Über einen die Erzeugung und Verwandlung des Lichtes betreffenden heuristischen Gesichtspunkt. *Annalen der Physik*, *322*(6), 132-148.

England, D. (2012). Digital art and interaction: lessons in collaboration. In *CHI'12 Extended Abstracts on Human Factors in Computing Systems* (pp. 703-712). ACM.

Erkut, C., Jylhä, A., & Disçioglu, R. (2011). A structured design and evaluation model with application to rhythmic interaction displays. In *Proc. Intl. Conf. New Interfaces for Musical Expression (NIME'11)* (pp. 477-480).

Fabiani, M., Dubus, G., & Bresin, R. (2011). MoodifierLive: Interactive and collaborative expressive music performance on mobile devices. *Proc. NIME 2011*, 116-119.

Fallman, D. (2003). Design-oriented human-computer interaction. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 225-232). ACM.

Fels, S., & Mase, K. (1999). Iamascope: A graphical musical instrument. *Computers & Graphics*, *23*(2), 277-286.

Friberg, A., Bresin, R., & Sundberg, J. (2006). Overview of the KTH rule system for musical performance. Advances in Cognitive Psychology, 2(2-3), 145-161.

Fritz, T., Jentschke, S., Gosselin, N., Sammler, D., Peretz, I., Turner, R., & Koelsch, S. (2009). Universal recognition of three basic emotions in music.*Current biology*, *19*(7), 573-576.

Gabrielsson, A. (1995). Expressive Intention and Performance. In R. Steinberg, *Music and the Mind Machine: The Psychophysiology and the Psychopathology of the Sense of Music*. Berlin: Springer Verlag, 35–47.

Gabrielsson, A., & Juslin, P. N. (1996). Emotional expression in music performance: between the performer's intention and the listener's experience. Psychology of Music 24(1), 68–91.
Gabrielsson, A., & Lindström, E. (2010). The role of structure in the musical expression of emotions. In P. N. Juslin, & J. A. Sloboda, *Music and emotion: Theory and research*. Oxford University Press, 367-400.
Gagnon, L., & Peretz, I. (2003). Mode and tempo relative contributions to "happy-sad" judgements in equitone melodies. *Cognition & Emotion*, *17*(1), 25-40.

Gartland-Jones, A., & Copley, P. (2003). The suitability of genetic algorithms for musical composition. *Contemporary Music Review*, *22*(3), 43-55.

Gaver, W. W., Bowers, J., Boucher, A., Gellerson, H., Pennington, S., Schmidt, A., ... & Walker, B. (2004). The drift table: designing for ludic engagement. In *CHI'04 extended abstracts on Human factors in computing systems* (pp. 885-900). ACM.

Gaver, W. (2007). Cultural commentators: Non-native interpretations as resources for polyphonic assessment. *International journal of human-computer studies*, *65*(4), 292-305.

Gilman, B. I. (1892). Report on an experimental test of musical expressiveness. *The American Journal of Psychology*, 4(4), 558-576.

Gomez, P., & Danuser, B. (2007). Relationships between musical structure and psychophysiological measures of emotion. *Emotion*, *7*(2), 377.

Gundlach, R. H. (1935). Factors determining the characterization of musical phrases. *The American Journal of Psychology*, 624-643.

Gurevich, M., & Cavan Fyans, A. (2011). Digital Musical Interactions: Performer–system relationships and their perception by spectators. *Organised Sound*, *16*(02), 166-175.

Hall, E. T. (1973). *The silent language*. Anchor.

Halpern, A. R., Martin, J. S., & Reed, T. D. (2008). An ERP study of major-minor classification in melodies.

Hargreaves, D., & North, A. (2010). Experimental aesthetics and liking for music. In P. N. Juslin, & J. A. Sloboda, *Music and emotion: Theory and research*. Oxford University Press, 515-546.

Henz, M., Lauer, S., & Zimmermann, D. (1996). COMPOzE-intention-based music composition through constraint programming. In *Proc. of Tools with Artificial Intelligence*. IEEE. 118-121.

Hevner, K. (1935). The affective character of the major and minor modes in music. *The American Journal of Psychology*, 103-118.

Hevner, K. (1937). The affective value of pitch and tempo in music. *The American Journal of Psychology*, 621-630.

Hiller Jr, L. A., & Isaacson, L. M. (1957). Musical composition with a high speed digital computer. In *Audio Engineering Society Convention 9*. Audio Engineering Society.

Hodges, D. (2010). Psychophysiological measures. *Handbook of music and emotion*, 279-312.

Hoeberechts, M., & Shantz, J. (2009). Realtime Emotional Adaptation in Automated Composition. *Audio Mostly*, 1-8.

Hornecker, E., & Buur, J. (2006). Getting a grip on tangible interaction: a framework on physical space and social interaction. In *Proceedings of the SIGCHI conference on Human Factors in computing systems* (pp. 437-446). ACM.

Höök, K., Sengers, P., & Andersson, G. (2003). Sense and sensibility: evaluation and interactive art. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 241-248). ACM.

Hunt, A., & Kirk, R. (2000). Mapping strategies for musical performance. *Trends in Gestural Control of Music*, *21*, 2000.

Ilie, G., & Thompson, W. (2006). A comparison of acoustic cues in music and speech for three dimensions of affect. *Music Perception*, 23(4), 319–330.

Jacob, B. L. (1996). Algorithmic composition as a model of creativity. *Organised Sound*, *1*(03), 157-165.

Johnston, A., Candy, L., & Edmonds, E. (2008). Designing and evaluating virtual musical instruments: facilitating conversational user interaction. *Design Studies*, *29*(6), 556-571.

Jordà, S. (2004). Digital instruments and players: part I---efficiency and apprenticeship. In *Proceedings of the 2004 conference on New interfaces for musical expression* (pp. 59-63). National University of Singapore.

Jordà, S., Geiger, G., Alonso, M., & Kaltenbrunner, M. (2007). The reacTable: exploring the synergy between live music performance and tabletop tangible interfaces. In *Proceedings of the 1st international conference on Tangible and embedded interaction* (pp. 139-146). ACM.

Juslin, P. N. (1997). Perceived emotional expression in synthesized performances of a short melody: Capturing the listener's judgment policy.*Musicae scientiae*, *1*(2), 225-256.

Juslin, P. N., & Sloboda, J. A. (2010). *Music and Emotion: Theory, Research, Applications.* Oxford University Press.

Juslin, P. N., & Timmers, R. (2010). Expression and communication of emotion in music performance. *Handbook of music and emotion: Theory, research, applications*, 453-489.

KaewTraKulPong, P., & Bowden, R. (2002). An improved adaptive background mixture model for real-time tracking with shadow detection. In *Video-Based Surveillance Systems* (pp. 135-144). Springer US.

Koelsch, S. (2011). Toward a neural basis of music perception – a review and updated model. *Frontiers in Psycology, 2*, 1-20.

Kwastek, K. (2013). *Aesthetics of Interaction in Digital Art*. MIT Press.

Legaspi, R., Hashimoto, Y., Moriyama, K., Kurihara, S., & Numao, M. (2007). Music compositional intelligence with an affective flavor. In *Proceedings of the 12th international conference on Intelligent user interfaces* (pp. 216-224). ACM.

Leman, M. (2008). *Embodied music cognition and mediation technology*. Mit Press.

Lerdahl, F., & Jackendoff, R. (1985). *A generative theory of tonal music*. MIT press.

Livingstone, S. R., Mühlberger, R., Brown, A. R., & Loch, A. (2007). Controlling musical emotionality: An affective computational architecture for influencing musical emotions. *Digital Creativity*, *18*(1), 43-53.

Livingstone, S. R., Muhlberger, R., Brown, A. R., & Thompson, W. F. (2010). Changing musical emotion: A computational rule system for modifying score and performance. *Computer Music Journal*, *34*(1), 41-64.

Machover, T. (1996). The Brain Opera and active music. *Catálogo Ars Electronica*.

Mancini, M., Castellano, G., Peters, C., & McOwan, P. W. (2011). Evaluating the communication of emotion via expressive gesture copying behaviour in an embodied humanoid agent. In *Affective Computing and Intelligent Interaction*(pp. 215-224). Springer Berlin Heidelberg.

Maslow, A. H. (1943). A theory of human motivation. *Psychological review*, *50*(4), 370.

Marquez-Borbon, A., Gurevich, M., Fyans, A. C., & Stapleton, P. (2011). Designing digital musical interactions in experimental contexts. *contexts*,*16*(20), 21.

Meyer, L. B. (2008). *Emotion and Meaning in Music.* University of Chicago Press.

Miranda, E. (2001). *Composing Music with Computers*. Butterworth-Heinemann.

Miranda, E. R., & Wanderley, M. (2006). New Digital Musical Instruments: Control And Interaction Beyond the Keyboard (Computer Music and Digital Audio Series), AR Editions. *Inc., Madison, WI*.

Miranda, E. R., & Al Biles, J. (Eds.). (2007). *Evolutionary computer music*. Springer.

Mitchell, M. (1996) *An Introduction to Genetic Algorithms*. Cambridge, MA: MIT Press.

Morreale, F., Masu, R., De Angeli, A., & Rota, P. (2013a). The music room. In *CHI'13 Extended Abstracts on Human Factors in Computing Systems* (pp. 3099-3102). ACM.

Morreale, F., Masu, R., De Angeli, A., & Fava, P. (2013b). The Effect of expertise in evaluating emotions in music. In *Proceedings of the 3rd International Conference on Music & Emotion (ICME3),* University of Jyväskylä, Department of Music.

Morreale, F., Masu, R., & De Angeli, A. (2013c). Robin: an algorithmic composer for interactive scenarios. *Proceedings of SMC*, *2013*, 10th.

Morreale, F., De Angeli, A., Masu, R., Rota, P., & Conci, N. (2014a). Collaborative creativity: The Music Room. *Personal and Ubiquitous Computing*, *18*(5), 1187-1199.

Morreale, F., De Angeli, A., & O'Modhrain, S. (2014b). Musical Interface Design: An Experience-oriented Framework. In *Proc. of NIME* (Vol. 14).

Morrison, A. J., Mitchell, P., & Brereton, M. (2007). The lens of ludic engagement: evaluating participation in interactive art installations. In*Proceedings of the 15th international conference on Multimedia* (pp. 509-512). ACM.

Nierhaus, G. (2009). *Algorithmic composition: paradigms of automated music generation*. Springer.

O'Neill, S. (1999). Flow theory and the development of musical performance skills. *Bulletin of the Council for Research in Music Education*, 129-134.

Oliveira, A. P., & Cardoso, A. (2010). A musical system for emotional expression. *Knowledge-Based Systems*, *23*(8), 901-913.

Overholt, D. (2009). The musical interface technology design space. *Organised Sound*, *14*(02), 217-226.

Pachet, F. (2003). The continuator: Musical interaction with style. *Journal of New Music Research*, *32*(3), 333-341.

Patel, A. D. (2010). *Music, language, and the brain*. Oxford university press.

Patten, J., Recht, B., & Ishii, H. (2002). Audiopad: a tag-based interface for musical performance. In *Proceedings of the 2002 conference on New interfaces for musical expression* (pp. 1-6). National University of Singapore.

Piston, W., & DeVoto, M.. (1978). *Harmony*. London: Gollancz.

Rigg, M. G. (1964). The mood effects

of music: A comparison of data from four investigators. *The journal of psychology*, *58*(2), 427-438.

Roads, C., & Strawn, J. (1985). *Foundations of computer music*. Massachusetts Institute of Technology.

Robazza, C., Macaluso, C., & D'Urso, V. (1994). Emotional reactions to music by gender, age, and expertise. *Perceptual and Motor skills*, *79*(2), 939-944.

Rota, P., Conci, N., & Sebe, N. (2012). Real time detection of social interactions in surveillance video. In *Computer vision–ECCV 2012. Workshops and demonstrations* (pp. 111-120). Springer Berlin Heidelberg.

Rowe, R. (1992). *Interactive music systems: machine listening and composing*. MIT press.

Ruismäki, H., Juvonen, A., & Lehtonen, K. (2013). The iPad and music in the new learning environment. *European Journal of Social & Behavioural Sciences, 6(3), 1048 - 1156.*

Russell, J. A. (1980). A circumplex model of affect. *Journal of personality and social psychology*, *39*(6), 1161.

Schlaug, G., Norton, A., Overy, K., & Winner, E. (2005). Effects of music training on the child's brain and cognitive development. *Annals of the New York Academy of Sciences*, *1060*(1), 219-230.

Schlosberg, H. (1952). The description of facial expression in terms of two dimensions. *Journal of Experimental Psychology, 4,* 229–237

Schlosberg, H. (1954). Three dimensions of emotions. *Psychological review*, 61, 81-88.

Schrödinger, E. (1926). An undulatory theory of the mechanics of atoms and molecules. *Physical Review*, *28*(6), 1049.

Schubert, E. (1999). Measuring emotion continuously: Validity and reliability of the two-dimensional emotion-space. *Australian Journal of Psychology*, *51*(3), 154-165.

Schubert, E. (2001). Continuous measurement of self-report emotional response to music.

Schubert, E. (2003). Update of the Hevner adjective checklist. *Perceptual and motor skills*, *96*(3c), 1117-1122.

Schubert, E. (2010). Continuous self-report methods. *Handbook of music and emotion: Theory, research, applications*, 223-253.

Schwartz, E., & Godfrey, D. (1993). *Music since 1945: issues, materials, and literature*. New York: Schirmer Books.

Sengers, P., & Gaver, B. (2006). Staying open to interpretation: engaging multiple meanings in design and evaluation. In *Proceedings of the 6th* *conference on Designing Interactive systems* (pp. 99-108). ACM.

Serra, X., Bresin, R., & Camurri, A. (2007). Sound and music computing: Challenges and strategies. *Journal of New Music Research*, *36*(3), 185-190.

Sharp, H., Rogers, Y., & Preece, J. (2007). Interaction design: beyond human-computer interaction. *2002*.

Sievers, B., Polansky, L., Casey, M., & Wheatley, T. (2013). Music and movement share a dynamic structure that supports universal expressions of emotion. *Proceedings of the National Academy of Sciences*, *110*(1), 70-75.

Simon, I., Morris, D., & Basu, S. (2008). MySong: automatic accompaniment generation for vocal melodies. In *Proc. of CHI 2008*. ACM. 724-734.

Steedman, M. J. (1984). A generative grammar for jazz chord sequences. *Music Perception*, 52-77.

Šimbelis, V., Lundström, A., Höök, K., Solsona, J., & Lewandowski, V. (2014). Metaphone: machine aesthetics meets interaction design. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*(pp. 1-10). ACM.

Temperley, D. (2004). *The cognition of basic musical structures*. MIT press, Cambridge.

Todd, P. M., & Werner, G. M. (1999). Frankensteinian methods for evolutionary music. *Musical networks: parallel distributed perception and performace*, 313-340.

Trochim, W. M. (1989). An introduction to concept mapping for planning and evaluation. *Evaluation and program planning*, *12*(1), 1-16.

Trost, W., Ethofer, T., Zentner, M., & Vuilleumier, P. (2012). Mapping aesthetic musical emotions in the brain. *Cerebral Cortex*, *22*(12), 2769-2783.

van Boerdonk, K., Tieben, R., Klooster, S., & van den Hoven, E. (2009). Contact through canvas: an entertaining encounter. *Personal and Ubiquitous Computing*, *13*(8), 551-567.

Varni, G., Mancini, M., Volpe, G., & Camurri, A. (2010). Sync'n'Move: social interaction based on music and gesture. In *User Centric Media* (pp. 31-38). Springer Berlin Heidelberg.

Vidyarthi, J., Riecke, B. E., & Gromala, D. (2012). Sonic Cradle: designing for an immersive experience of meditation by connecting respiration to music. In *Proceedings of the designing interactive systems conference* (pp. 408-417). ACM.

Västfjäll, D. (2010). Indirect perceptual, cognitive, and behavioural measures. *Handbook of music and emotion: Theory, research, applications*, 255-278.

Wallis, I., Ingalls, T., Campana, E., & Goodman, J. (2011). A rule-based generative music system controlled by desired valence and arousal. In *Proc. of*

8th international sound and music computing conference (SMC).

Wallis, I., Ingalls, T., Campana, E., & Vuong, C. (2013). Amateur musicians, long-term engagement, and HCI. In *Music and human-computer interaction* (pp. 49-66). Springer London.

Wanderley, M. M., & Orio, N. (2002). Evaluation of input devices for musical expression: Borrowing tools from hci. *Computer Music Journal*, *26*(3), 62-76.

Webster, G. D., & Weir, C. G. (2005). Emotional responses to music: Interactive effects of mode, texture, and tempo. *Motivation and Emotion*, *29*(1), 19-39.

Wiggins, G. A., Papadopoulos, G., Phon-Amnuaisuk, S., & Tuson, A. (1998). Evolutionary methods for musical composition. *International Journal of Computing Anticipatory Systems.*

Wilkie, K., Holland, S., & Mulholland, P. (2013). Towards a participatory approach for interaction design based on conceptual metaphor theory: A case study from music interaction. *In Music and Human-Computer Interaction* (pp. 259-270). Springer London.

Winkler, T. (2000). Audience participation and response in movement-sensing installations. *In Proc. of the International Computer Music Conference.*

Winner, E. (1982). Invented worlds: The psychology of the arts. *Harvard University Press.*

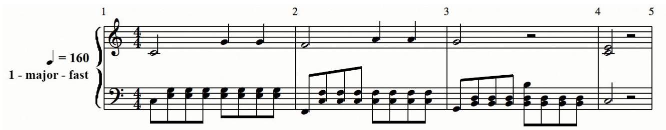
Zatorre, R. (2005). Music, the food of neuroscience?. *Nature*, *434*(7031), 312-315.

Zatorre, R. J., Chen, J. L., & Penhune, V. B. (2007). When the brain plays music: auditory–motor interactions in music perception and production. *Nature Reviews Neuroscience*, *8*(7), 547-558.

Zentner, M., & Eerola, T. (2010). Self-report measures and models. *Handbook of music and emotion: Theory, research, applications*, 187-221.

Zimmerman, J., Forlizzi, J., & Evenson, S. (2007). Research through design as a method for interaction design research in HCI. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 493-502). ACM.

Appendix

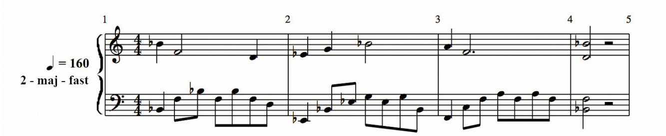


♩ = 160
1 - major - fast

♩ = 80
1 - major - slow

♩ = 160
1 - min - fast

♩ = 80
1 - min - slow



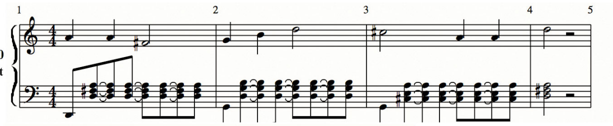
♩ = 160
2 - maj - fast

♩ = 80
2 - maj - slow

♩ = 160
2 - min - fast

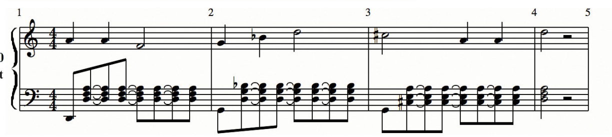
♩ = 80
2 - min - slow

♩ = 160
3 - maj - fast



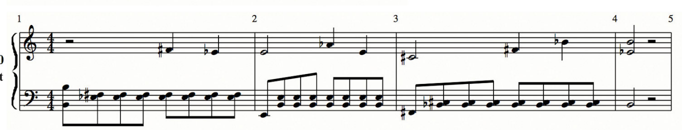
♩ = 80
3 - maj - slow

♩ = 160
3 - min - fast



♩ = 80
3 - min - slow

♩ = 160
4 - maj - fast

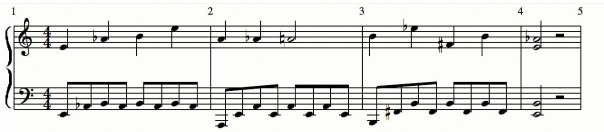


♩ = 80
4 - maj - slow

♩ = 160
4 - min - fast

♩ = 80
4 - min - slow

♩ = 160
5 - maj - fast

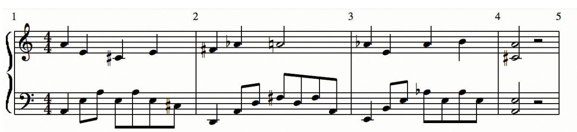


♩ = 80
5 - maj - slow

♩ = 160
5 - min - fast

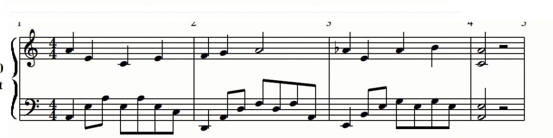
♩ = 80
5 - min - slow

♩ = 160
6 - maj - fast



♩ = 80
6 - maj - slow

♩ = 160
6 - min - fast

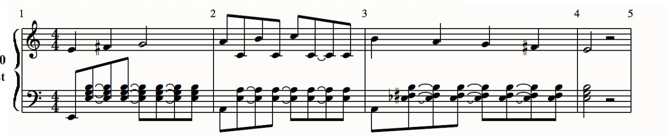


♩ = 80
6 - min - slow

♩ = 160
7 - maj - fast

♩ = 80
7 - maj - slow

♩ = 160
7 - min - fast



♩ = 80
7 - min - slow