Sede Amministrativa: Università degli Studi di Padova

Dipartimento di INGEGNERIA DELL'INFORMAZIONE (DEI)

SCUOLA DI DOTTORATO IN: INGEGNERIA DELL'INFORMAZIONE
XXVIII CICLO

Titolo della tesi:

# INTERACTIVE SPACES: MODELS FOR MOTION-BASED MUSIC APPLICATIONS


**Direttore della Scuola di Dottorato**: Ch.mo Prof. Matteo Bertocco
**Coordinatore di indirizzo**: Ch.mo Prof. Carlo Ferrari
**Supervisore**: Ch.mo Prof. Sergio Canazza

**Dottoranda**: Marcellina Mandanici

# Abstract

With the extensive utilization of touch screens, smartphones and various reactive surfaces, reality-based and intuitive interaction styles have now become customary. The employment of larger interactive areas, like floors or peripersonal three-dimensional spaces, further increase the reality-based interaction affordances, allowing full-body involvement and the development of a co-located, shared user experience. Embodied and spatial cognition play a fundamental role for the interaction in this kind of spaces, where users act in the reality with no device in the hands and obtain an audio and graphical output depending on their movements. Starting from the early experiments of Myron Krueger in 1971, responsive floors have been developed through various technologies including sensorized tiles and computer vision systems, to be employed in learning environments, entertainment, games and rehabilitation. Responsive floors allow the spatial representation of concepts and for this reason are suitable for immediate communication and engagement. As many musical features have meaningful spatial representations, they can easily be reproduced in the physical space through a conceptual blending approach and be made available to a great number of users. This is the key idea for the design of the original music applications presented in this thesis. The applications, devoted to music learning, production and active listening, introduce a novel creative approach to music, which can be further assumed as a general paradigm for the design of motion-based learning environments. Application assessment with upper elementary and high school students has proved that users engagement and bodily interaction have a high learning power, which can be a valid resource for deeper music knowledge and more creative learning processes. Although further interface tests showed that touch screen interaction performs better than full-body interaction, some important guidelines for the design of reactive floors applications have been obtained on the basis of these test results. Moreover, the conceptual framework developed for the design of music applications can represent a valid paradigm also in the general field of human-computer interaction.

# Sommario

Con l'utilizzo intensivo di touch screen, smartphones e varie superfici sensibili al tocco, stili di interazione intuitivi e ispirati alla realtà sono ormai diventati di uso comune. L'utilizzo di più estese superfici interattive quali pavimenti o spazi peripersonali tri-dimensionali, aumenta ulteriormente le possibilità offerte dall'interazione basata sulla realtà, coinvolgendo l'intero corpo e consentendo la condivisione dell'esperienza da parte di più utenti. La cognizione legata al corpo e allo spazio gioca un ruolo fondamentale per l'interazione in questi tipi di ambienti, dove l'utente agisce nella realtà senza dover tenere in mano alcun sensore, producendo un risultato sonoro e visivo in dipendenza dei suoi movimenti. Partendo dai primi esperimenti di Myron Krueger nel 1971, i pavimenti interattivi si sono sviluppati utilizzando varie tecnologie, dalle mattonelle sensorizzate a vari sistemi di computer vision, al fine di essere impiegati come ambienti per l'apprendimento e l'intrattenimento, per i giochi e per la riabilitazione. I pavimenti interattivi consentono la rappresentazione spaziale dei concetti e per questo motivo si prestano ad una comunicazione immediata e coinvolgente. Poichè molti elementi del linguaggio musicale sono espressi attraverso significative rappresentazioni spaziali, essi possono essere facilmente riprodotti nello spazio fisico utilizzando la teoria del "conceptual blending", ed essere così resi disponibili ad un numero assai esteso di utenti. Questo concetto è alla base della progettazione delle applicazioni musicali originali presentate in questa tesi. Le applicazioni, finalizzate all'apprendimento musicale, alla produzione e all'ascolto attivo, introducono un nuovo approccio creativo alla musica, che può essere considerato paradigmatico per la progettazione di ambienti educativi basati sul movimento. I test effettuati con alunni del secondo ciclo della scuola elementare e con studenti della scuola superiore hanno dimostrato che il coinvolgimento degli utenti e l'interazione "full-body" hanno una grande capacità di influenzare l'apprendimento, e che possono essere efficacemente utilizzati per un approfondimento della cultura musicale e per un approccio creativo nei processi formativi. Sebbene ulteriori test sull'interfaccia abbiano dimostrato che gli utenti preferiscono il touch screen all'interazione "full-body", in base a questi risultati sono state delineate alcune importanti linee guida per la futura progettazione di applicazioni su pavimenti interattivi. Inoltre, la struttura concettuale sviluppata per la progettazione di applicazioni musicali può rappresentare anche un valido paradigma nel campo dell'interazione uomo-macchina.

# Acknowledgements

First of all I want to thank my supervisor Sergio Canazza who gave me the possibility of beginning the research on interactive spaces at the laboratories of the Sound and Music Computing Group of the University of Padova. Thanks for his assistance, support and cooperation and, mainly, for considering my work worth of being shown in so many public events and demonstrations. Secondly I have to mention Antonio Rodà for his close presence and concrete help in many occurrences. His experience and knowledge was crucial in solving many problems, from scientific article writing to computer programming solutions. During the last three years the work at the laboratories of the SMC Group offered me many occasions of useful idea and experience exchanges. That's why I have to thank Giovanni De Poli, Federico Avanzini and Michele Geronazzo. In addition I owe also special thanks to Serena Zanolla who first introduced me to the fantastic experience of designing and testing interactive space learning environments. My work started from her "Stanza Logomotoria" and from the Zone Tracker application developed by Leonardo D'Amico. A great help, active cooperation and participation came from my colleagues Federico Altieri and Niccolò Pretto who have been involved at various levels in my research activities. And then I owe heartfelt thanks to all those school teachers and administrators who were eager to experiment the learning potentials of my music applications and devoted great efforts in organizing the assessment sessions at their institutions and in our laboratories. I want to mention in chronological order: the administrator of "Istituto Comprensivo di Passirano" (Brescia, Italy) and the teachers Paola Ballini, Marilena Abrami and Mariangela Agazzi; the administrator of the catholic institute "Barbarigo" (Padova, Italy) Cesare Contarini and the music technology teacher Giuseppe Viaro; the administrator of the catholic primary school "Gesù Maria" and the teachers Ilaria Zannoni and Laura Paliotto (Padova, Italy); the admnistrator of "Istituto Comprensivo Don Bosco" (Padova, Italy) Giovanni Zannoni and the teachers Elisabetta Pagani, Barbara Canacari and Donatella Orlando; the musicians Annalisa Meloni and Annabella Palleri and all the girls, boys and students who took part in the tests.

# Publications

*International Journals*

1. Mandanici, M., Rodà, & Canazza, S. (2016). The Harmonic Walk: An Interactive Physical Environment to Learn Tonal Melody Accompaniment. Advances in Multimedia, 2016, 1–16. doi:10.1155/2016/4027164

*International and National Conference Papers*

1. M. Mandanici, F. Altieri, A. Rodà and S. Canazza. Listen, Move, Learn: full body interactions in a music learning floor camera space. MET16 –Music, Education, Technology Proceedings London, 14–15 March 2016 (in press)

2. Mandanici, M. (2015, November). Interactive Spaces: Models and Algorithms for Reality-based Music Applications. In Proceedings of the 2015 International Conference on Interactive Tabletops & Surfaces (pp. 451-456). ACM.

3. M. Mandanici, A. Rodà and S. Canazza The "Harmonic Walk" and Enactive Knowledge: an Assessment Report In Proc. of the 12th Sound and Music Computing Conference Maynooth, August 2015

4. Mandanici, M., Rodà, A., & Canazza, S. (2015, March). A conceptual framework for motion based music applications. In Sonic Interactions for Virtual Environments (SIVE), 2015 IEEE 2nd VR Workshop on (pp. 1-5). IEEE.

5. Mandanici, M., Amico, L., Rodà, A., and Canazza, S. (2014). Conoscere l'armonia tonale nell'ambiente interattivo "Harmonic Walk". XX Colloquio di Informatica Musicale, Roma.

6. M. Mandanici, A. Rodà, and S. Canazza. The "Harmonic Walk": an interactive educational environment to discover musical chords. In Proc. ICMC-SMC Conference, 2014

7. Samory, M., Mandanici, M., Canazza, S., and Peserico, E. The Counterpoint Game: Rules, Constraints and Computational Spaces. In Proc. ICMC-SMC Conference, 2014

8. M. Mandanici and S. Canazza. The "Hand Composer": gesture-driven music composition machines. In Proc. 13th Conf. on Intelligent Autonomous Systems, pages 553–560, 2014.

*Submitted Journals and Conference Papers*

1. M. Mandanici, A. Rodà and S. Canazza. Bodily Interactions in Reality-based Music Environments. Submitted to Human Technologies, an Interdisciplinary Journal on Humans in ICT Environments, October 2015.

2. M. Mandanici, F. Altieri, N. Pretto, A. Rodà, S. Canazza. "Good or Bad?": an Augmented Reality Game to Engage Users in Active Listening Tasks. Submitted to the Journal of Immersive Education, April 2016

3. F. Altieri, M. Mandanici, S. Canazza, A. Rodà, E. Menegatti, M. Munaro "Good or Bad?": Potenziare l'ascolto attivo mediante spazi sonori interattivi. Submitted to XXI Colloquio di Informatica Musicale, May 2016.

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

*....Research should anticipate future practicality and not be bound by the constraints of the present. (Myron W. Krueger)*

## 1.1 What are Interactive Spaces?

Interactive spaces are two-dimensional surfaces or three-dimensional regions positioned in the range of some sensors or web of connected electronic devices, capable of detecting the presence, position or gestures of one or more users. Two-dimensional surfaces may be floors, boards, or walls; three-dimensional regions may be positioned everywhere and may be user-or sensor-centered. In any case, two- and three-dimensional spaces belong both to everyday life and, consequently, offer to the users immediacy of interaction and great feedback effectiveness. The applications built upon these principles bring in also another important quality: they have the power of mixing the reality of the actual space with the richness and variety of digital spaces, connecting thus the physical with the virtual world. Interactive spaces may occur in many dimensions and configurations and may be employed in various ways. The widespread utilization of cameras, camera software systems and motion tracking devices like Kinect, Kinect v2[1] and Leap Motion[2], have brought to the attention of a continously growing audience of researchers, designers and practitioners the great potentialities of motion as a more direct and intuitive way to interact with the computer. From late 2009, the year of Kinect's first version commercial launch, many applications have been developed for avatar games and entertainment,[3] hologram drawings for virtual and augmented reality environments,[4] for robot control,[5] and content management.[6] Nearly in the same years, some interesting experiences where made with active floors embedded

[1]Kinect `https://www.microsoft.com/en-us/kinectforwindows/`
[2]Leap Motion `https://www.leapmotion.com/`
[3]Xbox `http://www.xbox.com/en-US/?xr=mebarnav`
[4]Olivier Kreylos Homepage `http://idav.ucdavis.edu/~okreylos/index.html`
[5]Hybrid Systems Laboratory `http://hybrid.eecs.berkeley.edu`
[6]Evoluce `http://www.evoluce.com/de/index.php25.`

in schools or social spaces,[7] or used for active listening and music production,[8] for learning[9] and for entertainment.[10] It seems that the windows-icon-menu-pointer human-computer interaction paradigm is no longer sufficient to follow the development of the digital life of an ever growing number of users which, for the greatest part, employ a mobile device to interact with digital content and with their social community (van Dam, 1997). As a consequence, the reality-based actions connected to the widespread use of mobiles, like tapping, stretching and swiping, have become extremely common among users who probably are now ready to go further and demand new and larger physical places for digital social sharing. On the other hand it is also logical to expect that the hours and hours spent in pointing, typing and selecting, will in the long run claim a physical counterbalance. This is probably the reason why the research on motion tracking devices and on interactive spaces began, and why it is likely to become more and more important in the future.

## 1.2 Motivation and Research Approach

The focus of the present thesis is the study of active floors and of their affordances as music learning environments. Three-dimensional peripersonal spaces are also employed in two applications (Kinect Conductor and Hand Composer), due to their compliance to the model of the musical conductor. My research starts from the didactic experience developed with the Stanza Logomotoria, a multimodal active floor based on the "resonant memory" model,[11] which builds on the user's ability to memorize a sound linked to a precise spatial position. The acquired sound cognitive map drives the user to match a spoken text with a set of pre-recorded audio files, realizing an interactive sound story-telling reconstruction. This example involves spatial memory and full-body interaction and shows only one of the learning possibilities of active floor surfaces which, thanks to their dimension, allow experience sharing and social interaction. Many musical concepts, like harmony or melodic movements, have been historically depicted through very meaningful spatial representations, e.g. the Euler's tonnetz,[12] the Gregorian Neumatic Notation,[13] and Chironomy.[14] The very notion of musical instrument involves precise spatial rules and element arrangement, whose study has produced a large body of research about new music production interfaces in the field of the sound and music computing. Another interesting example is the music conductor model, where expressive performance informations are transmitted

[7] Interactive Spaces `http://www.interactivespaces.net/projects/index.php`

[8] Casa Paganini `http://www.infomus.org/research_ita.php`

[9] SMALLab `http://smallablearning.com/`

[10] Connected Worlds `http://design-io.com/projects/ConnectedWorlds/`

[11] Stanza Logomotoria `http://smc.dei.unipd.it/logomotoria.html`

[12] Literally "web of tones". The tonnetz appeared for the first time in a harmony treatise written by Euler in 1739. In the 19th century the tonnetz was reviewed and discussed by the German music theorist Hugo Riemann.

[13] The basic Gregorian chant note was represented by the "punctus", a square symbol put on the staff. Various combinations of "puncti" are called "neumes", which are customary spatial representations of the most used melodic patterns employed in the composition.

[14] Chironomy is a gestural system expressing melodic contour variations through hand movements. It was used to help Gregorian singers in tuning the right melodic pattern expressed by neumatic notation.

to performers through peripersonal space gestures. This led me to consider the possibility of employing physical surfaces or three-dimensional spaces to project geometrical representations of musical concepts, thus enabling users to enter and navigate conceptual maps conveying some musical knowledge. The experimental hypothesis upon which my work is based, is that in these environments implicit knowledge may emerge and drive the users to accomplish also very complex tasks, like melody harmonization. If this is true, enactive knowledge can be used to learn more in a shorter time, endowing motion-based learning environments with very powerful means of knowledge transmission. The aim of my work is to demonstrate how this cognitive mechanism can actively support music learning, offering new perspectives to music education and new possibilities for not professionals to have a deeper insight of musical structure.

## 1.3 Organization and Summary of Contents

An extensive review of the state of art of the technical development and employment of interactive spaces, with a particular regard to active floors and learning environments, introduces the topic of the thesis. Starting from Myron Krueger's early experiments in 1971, active floors develop hand in hand with computer vision techniques and nowadays benefit of software platforms capable of connecting the motion tracking data with a web of devices, like smartphones, tablets, wall screens, three-dimensional motion sensors and projectors, with the aim of creating entire responsive rooms or even larger spaces. From the very beginning of their history, the leading idea in the design of interactive spaces has been the need of sharing information to help human knowledge. Thus no wonder that one of the most important interactive spaces application field concerns the development of learning and entertainment environments, as shown in Chapter three. The affordances of bodily interactions inside an interactive space are analyzed in Chapter four, under the framework of what I consider the main themes of reality-based interaction: body, environment and social awareness and skills. Among the body skills, embodied cognition plays an important role as the leading factor in the determination of movement, whereas proprioception and the use of peripersonal space are fundamental in the musical practice also when playing a traditional instrument or when conducting a music ensemble. Users employ everyday life knowledge about their surroundings to build cognitive maps that allow navigation inside the interactive space. A cognitive map includes landmark and route knowledge; it is built with experience and then used to accomplish the application's required tasks. Acting in the same socio-spatial context with co-users or with bystanders, implies social interaction to be taken in account. Reinforcement, comparison and public monitoring are some of the elements useful to evaluate the psychological impact of social context on users experience and have to be considered as relevant factors in the application design. The link between the interactive landmarks spatial positioning and the conveyed musical meaning is one of the core topics of my thesis. This is the reason why the analysis of music spatial features and particularly of the representations of the harmonic space is so important. Moreover, in Chapter four also the characteristics of coordinated movement in a music environment are presented and discussed. Chapters five and six are devoted to the description and to the analysis of five case studies of motion-based music applications, subdivided in two groups: three applications employ reactive floors (Harmonic Walk, Jazz

Improvisation and "Good or Bad?"). The remaining two applications - Kinect Conductor and Hand Composer - employ a three-dimensional space. Harmonic Walk is a learning application for the study and practice of western tonal harmony; Jazz Improvisation is an active listening entertaining application, whereas "Good or Bad?" is a two player game which reaches a similar purpose engaging players through well defined gamification techniques (challenges, rewards, co-located and shared user experience). The three-dimensional space applications are both aimed at music production and expressive musical performance. They are based on the musical conductor model which is interpreted employing two different kinds of three-dimensional space: a body centered space in the Kinect Conductor and a sensor centered space in the Hand Composer. Various assessment sessions have been carried on for the Harmonic Walk application with high school and upper elementary children classes. The tests aimed at measuring the learning potential of the application, the performance differences between the active floor and the touch screen interface and how the spatial chord arrangement affects users movements and spatial memory. The assessment results are presented in Chapters seven and eight. They show that Harmonic Walk allows users to reach a very abstract and complex realm like tonal harmony in a quick and handy way, promoting the access of a potentially much larger number of people to the core of a fundamental musical structure. Thanks its direct and playful approach, the application proved to be very appealing and engaging for the children, both during the presentation sessions and during the tests. These two elements support the interest in proceeding with the research in this field, extending the didactic action to other domains like mathematics, science and foreign languages. Important guidelines emerging from assessment results point up the role of social interaction and immersiveness as fundamental elements in the design of reactive floors applications. Conclusion and further research development are presented in Chapter nine.

# Chapter 2

# Responsive Floors

In this Chapter the origins, aims and development of interactive spaces, and particularly of responsive floors, are thoroughly examined and discussed. Doug Engelbart's innovative ideas on human-computer interaction, Mark Weiser's vision of ubiquitous computing, and Bill Buxton's experiments at the Ontario Telepresence Project (Canada) and Xerox PARC (California, USA), are the core of the historical background of interactive spaces. The fundamental contributions offered by pioneers' early experiences further evolved in the research on media space technology and on ambient intelligence (AmI), which represent two different ways of interpreting the idea of interactive spaces. Starting from the work of a stunning forerunner like Myron Krueger, a series of responsive floors applications have been developed for learning, gaming, entertainment and other purposes. The timeline of active floors applications follows the development of motion tracking systems, which aim at supporting computer vision and motion data analysis for a more efficient and functional environment responsiveness. A short review of interactive spaces technologies and systems is provided at the end of the Chapter.

## 2.1 The origins of Responsive Floors

Responsive floors, as well as in general interactive spaces, are the result of the continous progress started with the work of Doug Engelbart[1] in the field of human-computer interaction. From 1962 Engelbart argued for the co-evolution of technology and human capabilities (Engelbart, 2001). An example of such co-evolution was the notion that a computer-mediated communication system that encouraged greater flexibility and interconnectivity would promote new office environments better supportive of collaborative work. Thus, research on comprehensive environments which could help workers in their everyday tasks led to the development of Engelbart's NLS (oN Line System), an intelligent working place supporting the most common activities, such as composing, studying, organizing, and modifying information (Engelbart, 1975). Another important experience in the field of computer supported workspaces was Richard Bolt's Spatial Data-Management system (Bolt, 1978) of the Architecture Machine Group at the Massachusetts Institute of Technology. The basis of the system is accessing a data item by going to where it is

[1] Doug Engelbart `http://dougengelbart.org/`

**Figure 2.1:** *A Media Space office set up at Xerox Palo Alto Research Center (PARC) with camera, microphone, video monitors and computer. Source* `http://people.cs.vt.edu/~srh/`

rather than referencing it by name, deploying spatial memory to retrieve documents represented on a two dimensional screen.

### 2.1.1 Media Space

In the eighties the availability of video technologies was growing and there were a range of explorations into the use of video for connecting people across distances. An example is Hole-in-Space (1980), a communication installation hosted in a public space. A real-time video/audio connection between Century City (in Los Angeles) and Lincoln Center (in New York City) was established in order to allow people to meet, see, and speak with each other as if they were in the same place. Further research at Xerox PARC led to the creation of the video-based system Media Spaces in 1993 (Bly et al., 1993). The system allowed people in remote offices, buildings, and even cities, to work together as if they were in the same architectural space. The technology supported a sense of shared presence and common social space which was independent of geographical location. Thus, from the very origins, the key concept of computer supported collaborative work was the creation of a virtual space which could exceed the boundaries of reality, allowing a better level of communication and knowledge sharing. This space is literally a media space, in that it is based on input devices like cameras and microphones connected through communication systems. The characteristics of media spaces represent an approach to design that is in contrast to personal computers, in which functionality is bundled into a single device, located at a single location, and operated by a single individual. On the contrary, systems should be embedded in the environment to explore the relationship between social function and architectural space. Bill Buxton's idea of UbiVid (Buxton, 1997) provides that rooms are equipped with a series of cameras and monitors to have visual data available at different sizes and locations. An UbiVid example is the Active Desk, developed at the Ontario Telepresence Project.[2] This is

[2]Ontario Telepresence Project `https://www.dgp.toronto.edu/tp/tphp.html`

a large electronic drafting table on which it is possible to interact with a stylus or a keyboard. Some cameras are mounted on each of the two upper corners of the desk thereby supporting multiple-way collaboration. The design being worked on is visible to all parties, and all parties can interact with it, such as pointing, making annotations, or adding to it. Thus "... The box into which we are designing our solutions is the room in which you work/play/learn, not a box that sits on your desk." (Buxton's design principle 4, *ibidem*). Moreover, Buxton's idea that "... Every device used for human-human interaction (cameras, microphones, etc.) are legitimate candidates for human-computer interaction..." (Buxton's design principle 5, *ibidem*), well establishes the concept upon which reactive environments are grounded. Thus, by mounting a video camera above the Active Desk, and feeding the video signal into an image processing system, one can track the position of the hands over the desk. The resulting signal enables the user to grasp computer-generated objects displayed on the desk's surface. As a matter of facts, current responsive floors, although employing different techniques, are no more than scaled camera spaces built on similar principles.

### 2.1.2 Ambient Intelligence

A complementary principle to Buxton's UbiVid is Mark Weiser's vision of ubiquitous computing. Weiser envisioned future environments where technology was embedded in every day objects but, at the same time, was not visible to the user (Weiser, 1991). To avoid the unnatural centrality of machines, ubiquitous computing should reside in the human world and pose no barriers to human interaction. From these ideas originates the concept of Ambient Intelligence (AmI) which is comprised of three main components: ubiquitous computing, ubiquitous communication, and user adaptive interfaces (Bick and Kummer, 2008). In general, ubiquitous computing technology is any computing technology that allows human computer interaction away from a single workstation, which includes pen-based technology, hand-held or portable devices, large-scale interactive screens, wireless networking infrastructure, and voice or vision technology. Moreover, visions of ambient intelligence and ubiquitous computing involve integrating tiny microelectronic processors and sensors into everyday objects in order to make them smart. Smart things can explore their environment, communicate with other smart things, and interact with humans, therefore helping users to cope with their tasks in new, intuitive ways (Bohn et al., 2005). Responsive floors may be considered as a part of ambient intelligence, since they are true, physic spaces which can be embedded in public places like classrooms, libraries, museums, as well as in private homes. Moreover, the unencumbered, completely natural interaction style proper of responsive floors fully resonates with Mark Weiser's idea of unobtrusive, simple technology. But, whereas ambient intelligence is based on the concept of ubiquitous and invisible computing, responsive floors, on the contrary, claim for a strong, localized presence, or even for a social role in the community (Petersen, 2005). Like a modern, technological agora,[3] a reactive floor has the power of attracting bystanders and to involve them in engaging social activities. Ambient intelligence embodies in some way a contrasting concept with respect to interactive spaces, mainly as far as concerns user interaction. In a smart environment the user has to know precisely where

[3]The agora was the custom public square in ancient Greece cities, the place of public discussion and decision.

and how to interact with the sensorized objects to obtain a feedack from the system. As Bellotti notices (Bellotti et al., 2002), there is an intrinsic communication problem concerning how this information must be delivered to the user, since one of the principles of ubiquitous computing is devices' invisibility. On the contrary, a responsive floor environment may be embedded in an everyday environment in a visible or invisible way. Anyhow, when interaction happens simply as a consequence of entering the active zone, it makes the user's experience more immersive and engaging.

## 2.2 The experiences of Myron Krueger

Myron Krueger is an american computer artist and researcher. He is considered the father of interactive art, the first promoter of the concept of "artifical reality" and of a free, unencumbered interaction style in responsive environments. Deeply dissatisfied with the restricted human-machine dialogue allowed by technology at that times (employing essentially keyboard and wand plus data tablet), he undertook a research for more interesting ways to interact with computers. His early projects, Glowflow (1969) and Metaplay (1971), introduced the concept of interactive art and provided the basic principles for next development (Krueger, 1977). Glowflow was a dark room installation where glowing lines of light defined illusory spaces. The lights and sounds produced by a Moog synthesizer responded to the footsteps of users when passing on a force sensitive pad, providing noticeable changes in the environment. In Metaplay, a real-time relationship between the user and the artist was established through a system of image transmission. The artist drawings were superimposed to the images of the audience in the installation room, providing a rich repertoire of interaction modalities between the artist and the installation users. These experiences led Krueger to consider interactive art as an autonomous new medium, completely different from traditional forms of art, like music, sculpture, and so on. In interactive art the quality of interaction is the focus of aesthetic evaluation, and not the audio-visual content, which rather must be chosen in order to convey a variety of conceptual relationships. For this purpose it is essential for users to be aware of how the environment responds to their actions, and for the computer to be able to perceive as much as possible of users behavior.

### 2.2.1 Psychic Space and VideoPlace

The next project, called Psychic Space, was designed in 1971 for an exhibition and is the first known reactive floor. A grid of 1,000 force sensors provided the computer with the position data of a single as well as of many users, whereas a rear projected screen displayed interactive graphics. The reactive floor of Psychic Space was used both as a musical instrument and as a visual experience. As a musical instrument, the floor reacted like a piano keyboard, with pitches' height depending on user position. But, if the user stopped moving for a while, the computer filled the gap repeating the more recent pitch sequence with height variations. The visual experience involved the concept of a mobile maze projected on the wall. As soon as the participant violated the boundaries of the maze, the computer answered stretching the line or moving the whole maze. These two examples demonstrate Krueger's basic idea of human interaction with an intelligent

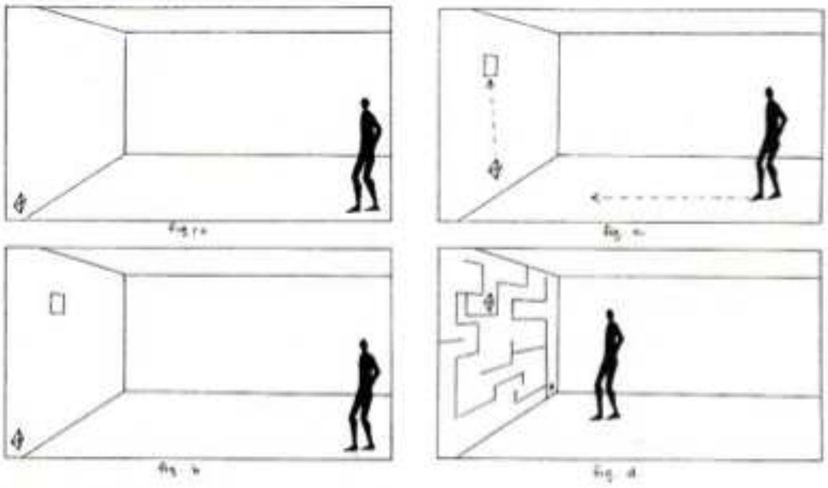


**Figure 2.2:** *The Psychic Space application with the mobile maze projected on the rear of a wall screen. Source* `https://www.studyblue.com`

machine and of how this interaction can be turned in an amusing, engaging and stimulating dialogue. Krueger conceived the human computer interaction in a responsive environment as an ".... encounter between individuals", where the participant communicates with motion and the entity responds with sound and graphycal output. Following this approach many other possibilities arise, like enhancing personal movement, peripersonal actions and capabilities, till considering the whole human body as an instrument. Moreover, the environment can be subdivided in many sub-environments inhabited by artificial characters defined by different sound and graphical output. The experience with the moving maze suggests that the relationship with the computer can be designed as a real game, where the user can make continuous discoveries and learning experiences. Further suggestions come from the VideoPlace responsive environment, designed in 1974. It consisted of two identical dark rooms which could be adjacent or far apart. In each room there was a camera and a rear-view projected screen which showed the image of the user himself in real size and the image of the participant in the other room. Thus users could interact between themselves or with graphycally represented objects. While the participants' bodies are bound to physical constraints, their images can be freely manipulated by the video processing algorithms, thus providing for the user a much richer and variable experience than reality itself (Krueger et al., 1985). VideoPlace further evolved in the VideoDesk application, where a ceiling-mounted camera looked down at the user's hands, allowing data manipulation using the hands image. This was employed also for teleconferencing, when two users at different location use their hands to point at shared documents on their computer screen (Krueger, 1993).

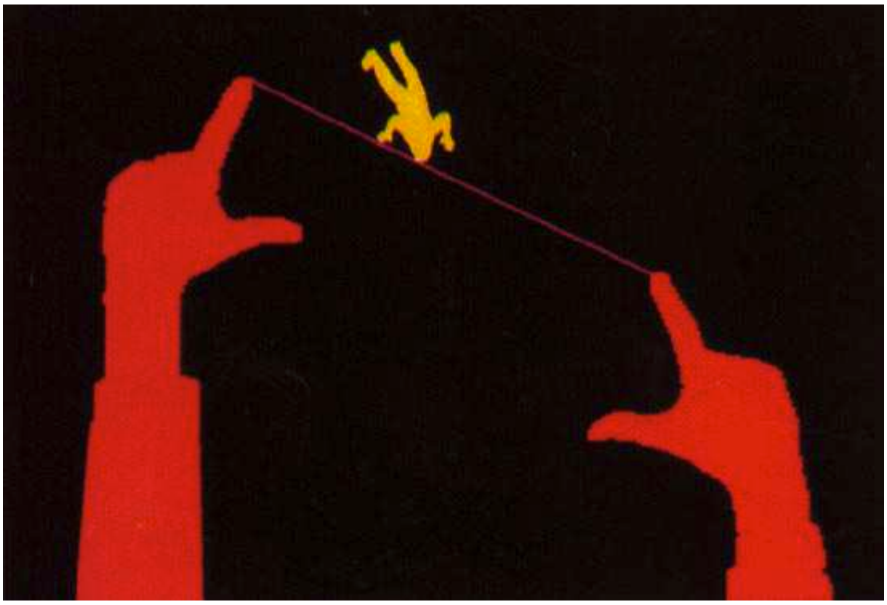


**Figure 2.3:** *Myron Krueger's VideoPlace application with image resize and manipulation. Source* `http://www.inventinginteractive.com`

### 2.2.2 Krueger's Vision

In spite of the technical limitations due mainly to the reduced computational power of machines at that time, it seems that Krueger was nevertheless able to hit many distinguishing points of human-computer interaction in responsive environments. He is the absolute precursor of interactive art and can be considered one of the pioneers of virtual and augmented reality research. Krueger fostered the importance of bodily intelligence as most important and popular with respect to abstract thinking. Bodily interaction in the physical world is much more immediate and simple to understand and represents a fundamental part of human knowledge. As a consequence, it must be regarded as an important alternative human-computer interaction modality. Despite the high number of reactive environments and applications produced, Krueger's approach to research was not limited to avant-garde engineering or design solutions. His main contribution is the observation and analysis of user's behavior and the interactive environment. The mechanisms of user's engagement, which is one of the conditions of entertainment, and the impact of interactive experience on user's cognition were the core of his considerations. This led him to formulate some interesting speculations about reactive environments' employment in the fields of education, psychology and pshycotherapy.

- **Education.** The high learning power of reactive environments is the use of physical movement as the interaction modality. Anticipating the theories of enactive and embodied learning, Krueger observes that in our schools learning is a sedentary activity imposed to naturally active children. He suggested that the exciting experience of intelligent interaction combined with physical movement could have a tremendous learning power. However, the design of learning environments should not be limited to automate traditional learning subjects, but should be aimed at changing the learning process dramatically. Since the environment can convey conceptual relationships which can be changed in a meaningful way, it should be possible to create simple interactions which enrich children's learning experience.

- **Psychology.** It is possible to deploy the environment's ability to monitor human behavior and to study human reactions. The use of responsive environments is particularly useful to study perception as a part of physical behavior; public situations provide a more spontaneous participants' behavior with respect to more controlled and artificial laboratory condition.

- **Psychotherapy.** The possibilities offered by a reactive environment to evoke and to expand physical interaction are very interesting for therapists. They can easily control the environment's characteristics and try to adapt them to the subject's needs.

The wide range of technological, cognitive and cultural consequences related to Krueger's work will not miss to affect future research, as witnessed by the ever growing interest in augmented and virtual reality. Anyhow, in spite of the great progress achieved by technology from Krueger's times, his approach and thoughts are far from being realized. As a matter of facts, the deep cultural changes he envisioned do not seem to be so easily at hand, nor the high learning power of reactive floors has yet been fully exploited by designers.

## 2.3 Responsive Floors Systems and Applications

In this Section an extensive review of the main responsive floors systems and applications is presented. The review is organized in the form of a timeline table reporting the year, the application's name, the research center where the project has been developed, the application's aim and the tracking system employed (see Table 2.1). The review includes only projects presented in a scientific paper, where the application's aim, and hardware and software tools are clearly described. Moreover, only responsive floors within the restricted area of a closed room are considered, with the aim of excluding other kinds of application employing big spaces interaction through RFID tags,[4] mobiles or other context-aware technologies. Of the thirty-two reviewed projects, eight are devoted to learning, six to gaming, five to music production and four to communication or social awareness. Other minor aims are entertainment, music active listening[5] and therapy.

[4]Radio frequency identification (RFID) uses electromagnetic field to identify and track tagged objects or persons in wide areas.

[5]Music active listening refers to the possibility for the user to make personal, interactive changes to what they listen employing sound synthesis algorithms or mixing techniques.

| *Year* | *Application* | *Research center/author* | *Aim* | *Tracking system/software* |
|---|---|---|---|---|
| 1991 | Intelligent Room | MIT AI Lab - Cambridge (USA) | Content management | Camera system/microphones |
| 1991 | Sound=Space | Rolf Gehlhaar | Music production | Ultrasonic echolocation system |
| 1994 | ALIVE | MIT Media Lab - Cambridge (USA) | Entertainment | ALIVE system |
| 1997 | Magic Carpet | MIT Media Lab - Cambridge (USA) | Music production | Sensorized floor/doppler radars |
| 1998 | Litefoot | University of Limerick (Ireland) | Feet movements detection | Sensorized floor |
| 1999 | Kidsroom | MIT Media Lab - Cambridge (USA) | Entertainment | Camera system/microphones |
| 2001 | Lumetila | VTT IT - Tampere (Finland) | Gaming | Sensorized floor |
| 2001 | Mediate | University of Portsmouth | Therapy | Sensorized floor/EyesWeb/microphones |
| 2005 | Color Race | Maersk Mc-Kinney Moller Institute for Production Technology - Odense (Denmark) | Game | Tangible tiles |
| 2005 | Pong | Maersk Mc-Kinney Moller Institute for Production Technology - Odense (Denmark) | Game | Tangible tiles |
| 2005 | Wicked Witch | Maersk Mc-Kinney Moller Institute for Production Technology - Odense (Denmark) | Game | Tangible tiles |
| 2006 | Orchestra Explorer | InfoMusLab - Genova (Italy) | Active listening | EyesWeb |
| 2007 | Mappe per Affetti Erranti | InfoMusLab - Genova (Italy) | Active listening | EyesWeb |
| 2007 | Lambent Reactive | USC School of Cinematic Arts Los Angeles (USA) | Music and visual production | Sensorized floor |
| 2007 | The Future Hybrid Library | Interactive Spaces (Denmark) | Communication | iFloor |
| 2007 | Pong | Interactive Spaces (Denmark) | Gaming | iGameFloor |
| 2007 | iFloorQuest | Interactive Spaces (Denmark) | Learning | iGameFloor |
| 2007 | Stepstone | Interactive Spaces (Denmark) | Learning/Therapy | iGameFloor |
| 2007 | Dashboard | Interactive Spaces (Denmark) | Content management | iGameFloor |
| 2007 | StorySurfer | Interactive Spaces (Denmark) | Communication | Camera system |
| 2008 | Bystander | University of Technology - Sidney (Australia) | Entertainment | Infrared cameras system |
| 2008 | CaDaReMi | Casa da Musica (Portugal) | Music production | Infrared camera system |
| 2009 | Springboard | School of Interactive Arts and Technology - Simon Fraser University (Canada) | Social awareness | Camera system? |
| 2009 | Stanza Logomotoria | SMC Group - University of Padova (Italy) | Learning | EyesWeb |
| 2010 | AmI Playfield | Institute of Computer Science Crete (Greece) | Learning | Cameras system |
| 2010 | SMALLab | Multimedia Arts Learning Lab University of Arizona (USA) | Learning | Camera and handheld devices system |
| 2011 | LPP | UCLA Los Angeles (USA) | Learning | Camera system |
| 2013 | Music Room | Experiential Music Lab University of Trento (Italy) | Music production | Camera system |
| 2014 | Kuniu Quantu | National Taiwan University | Gaming | Kinect system |
| 2014 | Harmonic Walk | SMC Group - University of Padova (Italy) | Learning | Zone Tracker |
| 2015 | Good or Bad? | SMC Group - University of Padova (Italy) | Active listening | Open PTrack |
| 2015 | STEP | Center for Research on Learning and Technology Indiana University (USA) | Learning | Open PTrack |

**Table 2.1:** *Timeline table of responsive floors applications and systems.*

### 2.3.1 Responsive Floors Technologies and Software Systems

From the technical point of view responsive floors are subdivided in two main groups. The first employs force sensors surfaces to detect users presence and position, whereas the other uses various camera-space systems and algorithms to detect not only presence and position but also users gestures and movement features. The short review provided below summarizes on technology progress and present state of art.

#### 2.3.1.1 Sensing floors systems

A grid of 1,000 force sensors has been the first responsive floor in Krueger's Psychic Space in 1971. Later on, Paradiso for Magic Carpet (Paradiso et al., 1997) employed a grid of piezoelectric wires for user position and a pair of Doppler radars to monitor user's movement to increase the sense of immersiveness. In the same years Litefoot, a system aiming at studying dance movements, was designed at the University of Limerick (Griffith and Fernström, 1998). In spite of force sensors good responsiveness, both systems suffered of signal confusion in time or space. To avoid this, a new collaborative project called Z-Tiles started in 2001 under the auspices of MediaLab Europe in Dublin. The design is grounded on a multiple sensors tiled surface employing blob detection algorithms for user's feet localization (Richardson et al., 2004). From 2000 a playware system employing tangible tiles was developed at the Technical University of Denmark. The tangible tiles are new play elements which function as building blocks by containing processing power, sensors, actuators, and communication capabilities. They provide the opportunities for creating new kinds of play and games. It is possible to have more than one game in the microcontroller and make different physical configurations of the tiles so the users can play different types of games (Lund et al., 2005). Contact sensing techniques, developed in 2010 allow to resolve the locations of the touch point, measuring the force applied on the interface. This method can be viewed as an efficient alternative to prior sensing techniques, as it requires far fewer sensors (Visell et al., 2010). A more recent floor prototype, GravitySpace employs a camera system located below the floor to sense pressure generated by people or room's furniture. While force sensors can track only what happens on the floor surface, this system can sense also above the floor e.g. allowing user's pose detection (Bränzel et al., 2013).

#### 2.3.1.2 Camera-space systems

The first camera-space systems were realized by Krueger's in the Videoplace application, where video projections were employed; or in the Metaplay installation, where live drawing interaction was performed by an artist on a data tablet and then projected on a wall screen. Starting from the ALIVE project (Maes, 1995), implemented at MIT Media Lab in 1994, many improvements were introduced, using computer vision algorithms for the tracking not only of the position of a person but also of a person's hands and head. Kidsroom (Bobick et al., 1999) further improved this technique, providing to recognize by the story a dozen of simple individual and group actions in specific contexts. The most popular and simple way to create a responsive floor is to place the floor area under the range of a ceiling-mounted camera at a convenient height, and to employ video streams data to feed real-time computer vision algorithms. In order to manage the noisy camera data, efficient algorithms must be implemented for data stream filtering and further processing. An important drawback of such systems is the dependence on environmental light condition which can alter the camera input and prevent efficient algorithm working. The Danish centre Interactive Spaces has employed two different responsive floor systems. One is the iFloor system which is and interactive floor which receives messages from SMS or emails. The floor interaction is based on a ceiling-mounted camera which tracks the people inside one meter band at the rim of the surface. People's position and movement are interpreted as magnetic forces attracting a cursor with its home position at the center of the floor display. The force is proportional to the size of the shadow blob generated by a person moving under the projector. iFloor maintains precise tracking of up to 10 people at one time in a 4x5 meters rectangle. The coordinates of the tracked persons are employed to calculate the movement of the cursor for application's content selection (Krogh et al., 2004). But, this technology is not enough efficient when a higher control level of many users movements is required. Thus, an interactive glassy floor platform called iGameFloor was designed and placed in a school department at Aarhus (Denmark). The system is based on four cameras put in a square box hollowed below the floor level. The users limbs can be tracked using webcams data to recognize the contact points created with the surface under suitable light conditions. It is thus possible to hit a button in an application even though other users are standing close (Grønbæk et al., 2007).

#### 2.3.1.3 The Zone Tracker application

A much simpler system for button function activation is the Zone Tracker application, implemented at the University of Padova (Amico, 2012). The video analysis algorithm analyzes the input images in three steps:

1. *Background subtraction.* The background is subtracted following the averaging background method proposed by (Jabri et al., 2000). The method used is based on statistical modeling of the background for each channel color and detection with adaptive threshold. The background model is obtained by calculating for each channel the average and the standard deviation of each pixel in the course of a certain number of frames acquired from static shots of the background. After building a statistical model of the background, the new frames are compared with it, thus producing what is called a confiance image, that is a grayscale image, where the value of each pixel represents the probability that it forms part of a region of the image not belonging to the background.

2. *Morphological image processing.* The images are processed by means of morphological transformations (Vincent, 1994). Dilation and erosion are the two basic mathematical operations which correspond respectively to the expansion and thinning of the black region. A combination of the two leads to opening and closure morphological operations, aiming at defining well shaped silhouettes blobs.

3. *Blob tracking.* The blobs moves are tracked and the two-dimensional barycenter of each blob is calculated.

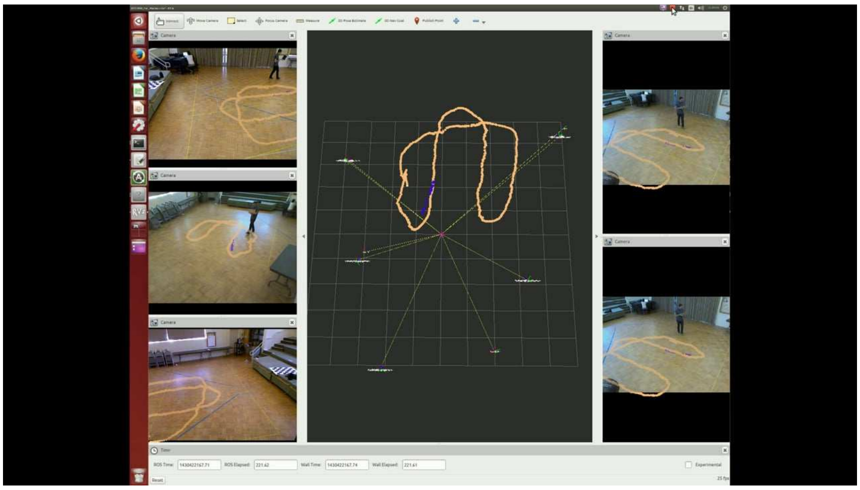


**Figure 2.4:** *Users tracking results obtained with a six sensors OpenPTrack system. Source* `https://vimeo.com/channels/864928`

The button function is activated by comparing the blob position with a customizable mask which partitions the floor surface for content synchronisation. Also if the application allows the tracking of more than one users, when people are close the two blobs may merge, thus generating some confusion in the tracking.

#### 2.3.1.4 The OpenPTrack project

Launched in 2013, the open source software OpenPTrack[6] is a scalable, multicamera solution for person tracking. OpenPTrack aims at supporting creative coders in the arts, culture, and educational sectors who wish to experiment with real-time person tracking as an input for their applications. The system allows to use a network of imagers to track the moving centroids (center of mass) of people within a defined area. It performs users detection in the machines connected to each sensor, whereas tracking is executed by a single node receiving data from all the network. These data can be incorporated into creative coding tools like Max/MSP, Processing, as well as a variety of other software languages and environments. (Munaro et al., 2016). It is very efficient for users tracking in every light condition and allows many users to be followed without any problem of occlusion.

### 2.3.2 Software Packages and Systems

One important motion analysis software tool is the EyesWeb[7] platform for the design and development of real-time motion-based applications. Designed and developed by InfoMus Lab since

[6] OpenPTrack `https://github.com/OpenPTrack/open_ptrack`
[7] EyesWeb `http://www.infomus.org/eyesweb_ita.php`

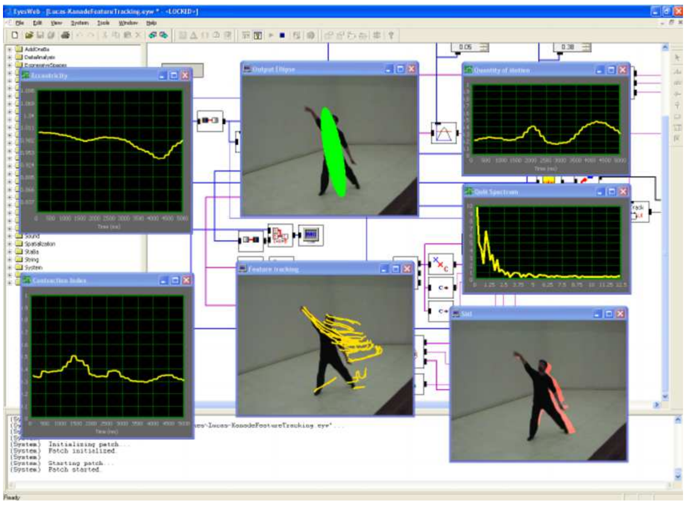


**Figure 2.5:** *An EyesWeb application extracting motion cues (Quantity of Motion, Contraction Index, and Kinematics cues). Source Camurri et al. (2003).*

1997, EyesWeb supports various standards and important functionalities like object Identification, segmentation and recognition, face recognition, gesture recognition and motion tracking. Moreover the EyesWeb Motion Analysis Library applies various signal processing techniques to extract expressive cues from human full-body movement (Camurri et al., 2004). The quantity of motion (QoM) computes silhouette motion images and delivers information about the variation of silhouette shape and position; silhouette shape and orientation computes the orientation axes on an ellipse that represents an approximation of the body; contraction index (CI) computes a measure of how the body employs the peripersonal space, by calculating the ratio between the body bounding region and the silhouette. These features represent important data which can be used for motion-related sound production, as in the Orchestra Explorer, where deictic gestures are employed to explore the sounds of the various instruments of the orchestra (Camurri et al., 2007).

#### 2.3.2.1 Interactive Spaces

The Google software platform Interactive Spaces[8] started in 2011. Written in the Java programming language[9] to guarantee the maximum of portability, the platform provides a collection of live activities as event producers and consumers in the physical space. The basic idea is to make the system easy to configure and to understand, and to endow it of all sort of functionality. It employs the ROS[10] as the main communication protocol and provides tools for web app and

[8]Interactive Spaces `https://code.google.com/archive/p/interactive-spaces/source/default/source`

[9]Java `https://www.oracle.com/java/index.html`

[10]ROS `http://wiki.ros.org/`

sockets availability, serial communication, Bluetooth communication,[11] XBee[12] and Android devices access for activity control.

### 2.3.3 The State of Art of Responsive Floors Systems

As made clear by the above review, systems evolution and usability is much more a question of software improvement and functionality than hardware quality. This is the reason why software tools are very important to achieve a high degree of responsiveness in interactive spaces. As far as concerns the performance characteristics offered by the sensing floors systems compared to the camera-space systems, some considerations can be made. Sensing floors systems allow for the use of simpler recognition algorithms and are less susceptible to occlusion. Nevertheless, they are demanding in terms of costs and are generally less portable and safe. Moreover, they offer limited flexibility and don't allow the tracking of users gestures. The above presented GravitySpace seems to offer some additional functionality in this direction, but is heavily demanding in terms of environmental requirements. On the contrary, camera-space systems allow a greater adaptability and offer important functions for the detection of users movements and movement features extraction. However, they are less robust in the tracking and are light condition sensitive. Occlusion and efficient multi-user tracking is available, but at the cost of heavy architectural interventions, as in the case of the Danish system iGamefloor. The OpenPTrack system offers promising advantages in terms of reliability and extensibility of the active area, at the only cost of heavy hardware requirements. Commercial systems like MotionMagix[13] or "GroundFX Interactive Floor Projection System" from GestureTek[14] offer all-in-one hardware set up solutions, but for a very restricted interactive area. Thus, the optimal compromise for interactive floors systems offering robustness in the tracking, a variety of motion features detection, portability, modular active area solutions and low environmental and hardware requirements, seems still far away.

[11]Bluetooth https://www.bluetooth.com/
[12]XBee http://www.digi.com/lp/xbee
[13]MotionMagix http://www.motionmagix.com/
[14]GestureTek http://www.gesturetek.com/gesturefx/productsolutions_groundfx.php

# Chapter 3

# Responsive Floors for Learning

The majority of the applications reviewed in Table 2.1 are devoted to learning. As predicted by Krueger, the peculiarities of interactive spaces play an important role in learning, thanks to the full-body interaction and to the immediacy of the audio and graphical feedback provided. The focus of this Chapter is on the employment of responsive floors in learning applications and on their educational design, that is the relationship between the application's content and the interaction required to the user. Many themes, different approaches and design principles emerge from the review of six responsive floors learning systems. Content analysis with respect to action, interaction modality and game tipology is presented and discussed.

## 3.1 Review of Responsive Floors Learning Systems

In this Section a review of six different systems and applications is presented. For each system is reported a brief overall presentation, system architecture description, assessment approach and results. The six systems are Wizefloor (Denmark), SMALLab (USA), Stanza Logomotoria (Italy), AmI Playfield (Crete), LPP (USA) and STEP (USA).

### 3.1.1 Wizefloor

The Danish research centre InteractiveSpaces was founded in 2002 by Kaj Grønbaek of the Department of Computer Science of the University of Aarhus with the aim of developing new ideas and concepts for interactive spaces. The main idea is to foster community interaction among co-located users through embedding interactive spaces in public spaces like schools, museums or libraries. The project's multidisciplinary approach brings together architecture, engineering and computer science and applies a participatory design method to better understand and satisfy the needs of users. Two responsive floor systems, iFloor and iGameFloor, were produced to support collaborative learning games, sharing knowledge applications and scenario simulation (Grønbæk et al., 2007). Collaborative learning games happen on a common, socially shared surface (like a traditional board) through direct bodily interaction. The research at Interactive Spaces centre

**Figure 3.1:** *A Wizefloor application. Source* `http://dirkovmolbaek.com/?cat=6`

led to the creation of Wizefloor, a product now commercialized by a partner of the centre, the Alexandra Institute.[1]

- **System.** Wizefloor comes in an unique ceiling-mounted hardware unit including camera, projector and loudspeakers, plus a miniPC, wireless keyboard and integrated pad. The system provides an interactive area of 2x3 meters with a precise tracking of multiple users, in every light condition but direct sunlight. An internet access is required to download the chosen game from a dedicated server and to record the game's results for group competition or teacher assessment. Game management and settings are available also through IOS applications, allowing a high level of content and fruition modalities customization.

- **Assessment.** Assessment is provided for two experimental set ups: the iFloor and the iGameFloor, both described in Section 2.3.1.2. The iFloor system has been employed in the Future Hybrid Library project, with the aim of enhancing the social aspects of the library. The iGameFloor has been evaluated in the framework of the Stepstone game, where children have to jump on a virtual stone to select a right answer.

- **Results.** As a part of a qualitative iFloor evaluation, a school class was observed in a one hour quiz game. The experience revealed the application's limits in managing a great number of participants, which led the cursor to be blocked by too many inputs. Children were quick in understanding the affordances of the system, which very soon was forced to load

[1]Wizefloor `https://www.wizefloor.com/`

**Figure 3.2:** *A group of students in the SMALLab interactive floor with the Scalebo application running. Source* `http://smallablearning.com/`

personal and mates questions and answers, unrelated to the library's information context. Also the Stepstone proved to be intuitive to use for the children. It enhances collaborative behavior in a funny and engaging way. Technical problems emerged in the individual user's blob identification, as in some conditions it was difficult to avoid confusion in multi-user tracking.

### 3.1.2 SMALLab

Based on a technology platform developed by David Birchfield at the Arizona State University (2008), SMALLab (Situated Multimedia Art Learning Lab) uses motion-capture cameras, short-throw projectors and wireless controllers to immerse players in dynamic game-based scenarios, where they can interact with each other and digital game elements in real time in shared physical space. The SMALLab Learning is the commercial company for SMALLab from 2010.

- **System.** SMALLab is a square interactive floor approximately of 4.5x4.5 meters with a set of wireless input devices (e.g., Gamepad, Wii Remote), a top-mounted video projector providing dynamic visual feedback on the floor, and four audio speakers for surround-sound feedback. The system uses a networked computing cluster with custom software developed by the Arizona State University research team.

- **Assessment.** System evaluation was carried in the framework of two learning application's scenarios: the Layer Cake for the study of geological evolution and the Physics-Spring Pendulum for the study of coupled harmonic oscillators (Johnson-Glenberg et al., 2011). Both scenarios depict an interactive visual scene related to the topic and allow physical interaction from users. In the Layer Cake application students' interaction aims at placing fossils and sediments in the right place and geological ages through embodied interaction. In the Physics-Spring Pendulum students had to oscillate particles using spring pendulum motion.

- **Results.** The Layer Cake experimental hypothesis aimed at demonstrating that students in the SMALLab condition would learn more because their learning will be embodied and multimodal. The subjects where all $9^{th}$ graders (14-15 years old). Group 1 was composed of 34 subjects instructed first in the SMALLab condition and after in the regular class. Group 2 was composed of 37 students instructed first in the regular class and after in the SMALLab condition. A total gain effect size of 19.42 was achieved by group 1, where a total gain effect size of -1.56 was achieved by group 2, thus demonstrating that receiving regular instruction before exposure to SMALLab significantly affected overall learning of content, at least for this sample. The Physics-Spring Pendulum assessment aimed at measuring the impact of an embodied learning experience on knowledge gains. Two groups (15 and 16 participants from the undergraduate Psychology subject pool at Arizona State University) in two different conditions were tested: the first group was tested in the SMALLab physical condition, where swinging gesture would compress or stretch the spring more than a smaller movement; the second group in the desktop condition, where the same changes happened moving a mouse. Results show a general knowledge gain for both groups, with no significant surplus gain for the SMALLab group, as expected.

### 3.1.3 Stanza Logomotoria

The Stanza Logomotoria was implemented at the University of Padova (Sound and Music Computing Group) in 2008 with the aim of promoting learning activities which employ an augmented reality physical environment. The main Stanza Logomotoria application is the Resonant Memory, where the space is subdivided in nine areas: eight peripheral and one central. The user's presence within a specific area triggers the audio reproduction of the corresponding content. Noises, music and environmental sounds are synchronized to the peripheral areas, whereas the central area is synchronized with a storytelling. The child first explores the sounds and then tries to match them to the story. The use of body movement associated to spatial localization helps sound memorization and strongly enhances listening abilities, as the application's name well explains. The Resonant Memory has been used in elementary schools to compose a sound storytelling or to learn a foreign language.

- **System.** The Stanza Logo-Motoria is a modular system composed by an empty room equipped with a ceiling mounted webcam and loudspeakers. The user's body movements are processed by a software module developed in the EyesWeb XMI environment,[2] which performs the analysis of the input video stream, the control of the interaction logic, and the audio rendering task.

- **Assessment.** An assessment protocol was established to verify if pupils who use the Stanza Logo-Motoria as a listening tool in english lessons improve more significantly in word

[2]EyesWeb XMI `http://www.infomus.org/eyesweb_ita.php`

**Figure 3.3:** *The nine coloured areas representing the nine Resonant Memory application's zones (eight peripheral and one central). The grey diamonds are the blind zones left empty to avoid zone confusion. Source Camurri et al. (2010).*

recognition and oral language comprehension than those who perform listening by means of a CD player and two loudspeakers in the language laboratory. The use of the Resonant Memory application was employed in the experimental group (sixteen pupils) while the control group employed traditional language laboratory methods (twelve pupils). All the pupils were aged 9.

- **Results.** The assessment test's results show that pupils trained with the Resonant Memory application are more likely to comprehend and remember what they listen to compared to pupils using a traditional language laboratory (Zanolla et al., 2013). After the experimental treatment, the average group scores significantly differed: 89% (average experimental group score) and 74% (average control group score); the average experimental group score increased largely compared to that of the control group.

### 3.1.4 AmI Playfield

AmI Playfield has been designed starting in 2009 in the framework of the AmI (Ambient Intelligence) Programme at the Foundation of Research and Technology, Hellas - FORTH in Heraklion (Crete). The programme aims at creating experimental AmI spaces for art and culture, commerce, healthcare, home, learning, education and workplaces. "AmI Playfield" combines location awareness with multimodality to promote entertainment and learning, enhancing physical play.

- **System.** The playfield consists of a 4x4 $m^2$ carpet, shaped as a grid of game positions by several equally-distanced plastic stripes, of a parallel dual back-projection display, of an information kiosk used to host a remote game manager, of several Wi-Fi enabled mobile phones used as additional game controllers and of a surround speaker system. Moreover

**Figure 3.4:** *The Ambiente Intelligence environment with the Apple Hunt application running. Source* `http://www.ics.forth.gr/ami/projects/`

AmI Playfield is endowed with a performance measurement system which stores all player actions for further processing. Depending on the learning scenario, physical and digital marks are assigned to each game position that may vary from arithmetic or alphabetic to custom. Various graphical user interfaces convey the image of the game's virtual world to the dual back-projection display, in order to aid in the players decisions and activity.

- **Assessment.** System assessment has been done using the Apple Hunt learning application, a game which addresses fundamental arithmetic operations and is oriented towards elementary school children. Apple Hunt defines various targets and obstacles, represented naturally by apple and trap dummies, which are randomly placed on the playfield, according to the system's output. The objective of the game is to pick as many apples as possible by the end of the game, avoiding the traps. The players collaborate freely in coordinating their actions to gather the most of the apples available on floor, while calculating simple arithmetic operations required for moving from one place to another. Running, yelling, jumping or even "cheating" with the help of the attendants are all part of the game, while traps impede in the players' efforts.

- **Results.** The assessment of the Apple Hunt game so far has involved nine children aged from seven to eleven years (Papagiannakis et al., 2013). The evaluation approach avoids deliberately application's efficiency assessment, focusing instead on fun, challenge, engagement, learnability, collaboration and kinesthetic activity. The main findings relating to fun include expressions of joy and laughter, enhanced by all participants' spontaneous will to continue playing at the end of the sessions. Negative expressions or boredom were not observed. Learning the game did not pose difficulties to the participants, whereas all participants experienced full engagement in the game.

### 3.1.5 Learning Physics through Play

The Learning Physics through Play Project (LPP) is based on the idea that young children can, under the right circumstances, learn more complicated ideas than we currently ask of them in

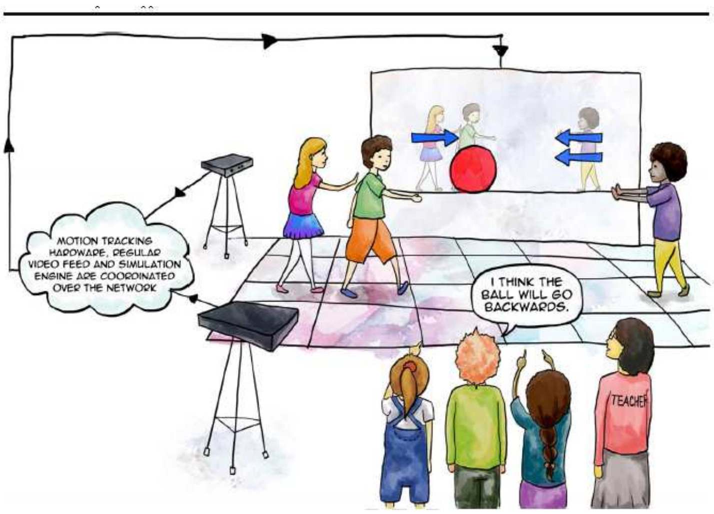



**Figure 3.5:** *A LLP application where students model a ball's behavior according to their own physical movements. Source Enyedy et al. (2015).*

early elementary science education. They can engage in productive inquiry, collect and analyze data, produce models, and learn complex concepts. The projects aims at these important results employing two design principles: the use of a dramatic, embodied play to model and support inquiry and the use of progressive symbolization to help students construct meaning (Enyedy et al., 2012). LLP bas been used in a series of scientific investigations of Newtonian force and motion. Students started the activity using ball-play skills to make a prediction. The technology translated the students' physical motion during play into a augmented-reality, computer animation displayed on the whiteboard and combined the students' motion with symbolic elements that marked important points in the embodied prediction. Through this game-like experience, LLP made it possible for 6-8 year-old students to interrogate their own understanding and explore these physics concepts.

- **System.** The system consist of a 3.5x3.5 meters carpet and of a whiteboard for LLP microworld projection. The LLP system is based on an augmented reality system that uses computer vision to track students physical actions and locations. The motion data are processed by a physics engine which generates responses based on students data.

- **Assessment.** The LLP was tested with two multi-age classrooms gathering students aged 6 - 8 years. Students were presented with objects (e.g., soccer balls, volleyballs, basketballs, and girls on skateboards), traveling across one or more surfaces (e.g., a wood floor, grassy lawn, and an ice skating rink), and asked to reason through different situations (e.g., racing, or a sequence of different surfaces and/or forces). For each scenario, students were asked to identify where the forces are, to describe qualitative and quantitative differences in applied forces on various surfaces, contexts, objects, and scenarios, and to make predictions about the resulting speed or direction of applied forces.

- **Results.** Descriptive statistics were obtained on performance on the pre-test and post-test items. For the 43 students, the average pre-test score was 5.42 (SD=1.38) out of a possible of 16 points. The average post-test score was 8.54 (SD=2.17). Post-test scores were significantly higher than the pre-test scores, and the effect size of the gain was large, d=1.99. These findings suggest that many of the students made significant progress in learning the content as measured by the pre- and post-test.

### 3.1.6 Science through Technology Enhanced Play

The STEP environment is a prosecution of the LLP project. It uses mixed reality tools to combine students' embodied play activities with the power of computer simulations and thus support them in reflecting on scientific content.

- **System.** The STEP environment employs the OpenPTrack system for accurate users motion tracking (see Section 2.3.1.4).

- **Assessment.** For the STEP environment assessment a scenario displaying the different states of water has been prepared. Students move around in the classroom space, play-acting how they believe water would behave in a range of circumstances such as a freezing cold day. They can see their activity projected into a computer simulation where an avatar of a water particle is displayed over the video display. This avatar includes a representation of important information necessary for the students' inquiry such as the temperature in the imagined space or the energy level of the individual particles. As the students role-play in the STEP environment, they make choices about how real particles might behave, and those choices are reflected within the projected simulation. This study was implemented with eighteen second-grade students (ages 7-8, twelve male, six female).

- **Results.** After engaging with the STEP activity, students were able to articulate accurate descriptions of particle behavior in different states of matter, as well as the relationship between energy, particle behavior and state changes. Student's answers in these domains improved significantly, from 8% to 56%, indicating that students generally provided more robust and normative accounts for how matter changes after engaging in the STEP activities.

## 3.2 Content Analysis of Reactive Floors Learning Games

Relying on information available on commercial systems public websites[3],[4] or from scientific papers content, Table 3.1 has been prepared with the aim of listing and of analyzing the contents of the games offered by the various systems. As information was not always complete or homogeneous, the table presents some missing points and, probably, also some mistake. Anyhow, it can be considered as a tentative approach to provide a comprehensive review aimed at

[3]WizeLearning `http://www.wizelearning.org/applications/`
[4]SMALLab Learning `http://smallablearning.com/learning-scenarios/`

giving a general idea about the learning possibilities of the various systems. The table groups the reviewed games by system. Action, interaction modality and typology are reported for each game. The action field describes what the user does to fulfill game's requirements; the interaction modality explains how the activity is embodied through player's movement; the typology tries to capture the game's purpose and meaning.

### 3.2.1 Game Actions

Game actions are connected to game's aim and functions and to the interaction logic. What follows is an annotated list of the most common actions observed through the games documentation analysis.

1. **Located selection.** The action of selecting a spatially located item is linked to the meaning of making a choice, through the activation of a specific button function. The choice is useful for many purposes, like indentifying a right answer for a quiz (Stepstone, Quiz), choosing a symbol (letter or number as in Balloons), removing an item (Scratch), making things do something (Flowers, Rubber ducks), displaying information (Slides, Poll, Character profile).

2. **Coordinated selection.** The combination of one or more selections allows for more complex functions, where coordinated selection is needed. Coordinated selection may happen or through multi-user interaction or through buttons with hierarchical functions. For example, when ordering symbols in the Floor Keyboard game, the enter function is operated by a dedicated button which is stepped on at the end of the required symbol's sequence; another example is the structured content of the Resonant Memory application, which is activated only after the exploration task has been accomplished. Coordinated selection is also involved in the aim of recognizing a relationship between items, and of connecting them (Paper lines, Categorize); or it can be employed with the aim of creating something depending on multi-points selection (Geometrix or Color Mixer).

3. **Walk.** The action of walking may be free like a wandering (Flowers) or target-directed (to the position of the virtual ball as in Football). It can be used to move along abstract representations like scaled values (Fraction Lab) o graphs (Constant Velocity). Users walk may be slow or fast (running). The kinematic data thus generated can be used to match body movement with the physics of an event (STEP), or to model or to predict a ball's behavior (LLP). User's walk can further be used to define a track. In the Apple Hunt game the track-finding is defined by the functions necessary to solve elementary mathematical problems.

4. **Body parts movement.** In some environments (SMALLab, Stanza Logomotoria) also hands, arms and shoulders movement can be tracked. Any body part movement can be used as a generic button function, allowing for example the progress of a story-telling operated also by impaired subjects (Fiaba Magica). Again, upper limbs kinematic data can be employed to model mechanic concepts manipulation (Le Tour de Force, Lifting Gears) or to provoke virtual earthquakes (Geology Layer Cake).

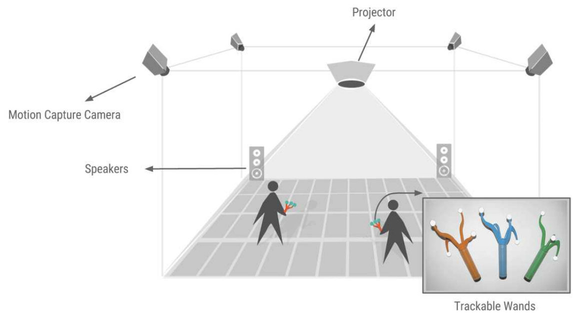



**Figure 3.6:** *The custom trackable wand with passive retro-flective markers in the SMALLab environment. Source* `http://www.kh-liu.com/smallab.html`

### 3.2.2 Interaction Modalities

The interaction modalities employed in responsive floors learning environments are connected to the motion tracking devices and techniques, as well as to the interaction logic involved in the application. Among the six reviewed systems, five employ only free, unencumbered bodily movement, whereas SMALLab combines computer keyboard or remote controller input with data coming from one or more custom trackable wands. The wands are endowed with passive retro-reflective markers which are recognized as a unique piece by the motion capture system. As shown in Section 3.2.1, input modalities of unencumbered body tracking are:

1. simple selection (e.g. an arm gesture);
2. localized selection (a step or jump on an interactive landmark);
3. coordinated selection (more than one selection coordinated in a hierarchical way);
4. continous data (the tracking of a walk or run, or of upper limbs movements).

On the other end, the SMALLab wand biases user's movements in many ways. First of all, it needs to be kept in user's hand and well far apart from the body to be seen by cameras. Moreover, it acts as the only transmission mean between user and responsive floor, channeling in this way all user's movements. In return, it can deliver a lot of interaction data that can be used in many ways:

1. for localized selection on the responsive floor, that can be obtained tapping at a targeted position or hovering over an image;
2. for the activation of the "drag and drop" function through a double tap;
3. for continous data generation obtained walking with the wand on the floor, or moving it along body height, swiping it for rotation and other kinematic data.

It is clear that the user's freedom when moving without any device in the hands is counterbalanced by reduced functionalities in the interaction. For instance, the "drag and drop" function is hard to implement when tracking only the user's blob, whereas coordinate selection is not so fluent as one might desire, as it is mediated by stiff programming constraints. The SMALLab environment has also another important feature that is the possibility of switching from one scenario to the other with computer keyboard or remote controller used by the teacher. This leads to the possibility of organizing complex learning scenarios in a progressive way, creating different learning levels inside the same application.

### 3.2.3 Game Typologies

From the overview of the six responsive floor systems considered in the present review, a total of eight game typologies have been identified. Game typology deals with application content and aim, and summarizes game cognitive goals and meaning.

1. **Sensory Environments.** A first remarkable group of applications are sensory environments, usually designed for pre-school children. The aim of this kind of application is to solicit an emotional response through remarkable environmental reaction (popping images, unexpected sounds, etc.). A number of similar interactive spaces have been implemented for therapy or rehabilitation purposes, including modular systems of interactive objects (Chung and Lai, 2002), multisensory interfaces environments (Gumtau et al., 2005) or combination of virtual worlds, large screens, mobile robots, and smart objects (Agosta et al., 2015). Sensory environments help to awaken emotions useful to combat isolation and to elicit positive behaviors.

2. **Impaired users games.** Games for impaired users may allow full and satisfactory interaction experiences to people suffering of many kinds of disabilities. Depending on the kind of disability, the most suitable interaction modalities can be chosen to fit user's needs. This game typology has only two examples in the present review: nevertheless, it is an interesting typology of applications, because it emphasizes responsive floors affordances and flexibility. Users can be engaged not only with the emotions deriving from the interaction but also in cognitive tasks like matching sound and images (Iversen et al., 2007) or creating and interactive story (Zanolla et al., 2013).

3. **Real games reproduction.** These applications are designed with the aim of reproducing already existing traditional games on a responsive floor. The benefits of playing real games in an augmented reality environment are to help users to learn how to interact on the floor and gain knowledge about the technology used. Employing already acquired sensorimotor schemas and game logic helps users to concentrate on the new environment and to discover how to exploit its potentialities.

4. **Memory games.** Memory games exploit located items spatial relationships to remember the place of the objects to be connected. This cognitive mechanism, known as the "Si-

monides effect”[5] has been employed in the already cited Spatial Data Management System (Bolt, 1978), a forerunner of the window and icon computer screen organization. It is of particular interest in responsive floors environments, where items have to be arranged on a floor surface following some organizing principle. As an example, in the Resonant Memory application users explore the environment and build a cognitive map of the sound locations. Afterwards, they employ spatial cognition to interpret the cognitive map to direct their movements to accomplish the application’s tasks (Camurri et al., 2010).

5. **Information display games.** Information display games employ bodily interaction to communicate content to others. The act of sharing information is enhanced by embodied interaction, which gives more strength to the message and empowers social interplay. Users motion kinematic qualities are perceived by participants as linked to the message content, endowing the communication with unexpected implicit meanings.

6. **Inquiry learning games.** Inquiry learning games is a very popular game category. In these games a learning scenario is proposed for user interaction, where the system responds with some changes depending on user’s actions and movements. The changes follow some non explicit rule (e.g. a physics law) which the user has to discover. Thus the free wandering exploration is the basic user behaviour in these environments. Inquiry learning derives from the “learning by doing” approach, which affirms that true knowledge may happen only after a meaningful experience and not only after passive abstract content transmission (Schön, 1992).

7. **Challenge games.** Challenge games are learning environments where the user is asked to answer a question or to solve a problem. Thus, not only the user is required to explore the environment and to understand how it works, but has to leverage on her/his skills, attention and knowledge. Embedding challenges in learning environments means to make the embodied experience more engaging for the user. A challenge implies also rewards and score assignment, which fosters individual competition and group cooperation. Moreover challenges in a learning environment may be used as a real assessment tools.

8. **Play-acting games.** Play-acting games are similar to inquiry learning games in many aspects. Also play-acting games involve a learning scenario representation as well as its exploration and discovery. What characterizes play-acting games is the type of interaction that depends on movement kinematic attributes. These attributes are employed to play-act events and situations of the physical world trying to embody some particular aspect. In play-acting games the users behave as if they were the event or the element at the core of the application (i.e., spinning arms in circle to increase or diminish a bike’s speed, or making slow movements to emulate freezing water particles). This is the rendering of abstract concepts through body movement features and represents the highest degree of embodied knowledge.

[5]The Greek poet Simonides (556-468 B.C.E.), inferred that persons desiring to train their memory must select places and store there content to be remembered, as the order of the places can recall the order of the things.

| *System* | *Game* | *Aim* | *Interaction modality* | *Typology* |
|---|---|---|---|---|
| Wizefloor | Floor menu | selection | step | information display |
| | Quiz | answer yes or no | step | challenge |
| | Stepstone | quiz solution | step | challenge |
| | Stepstone Ling Game | quiz solution | step | therapy |
| | Balloons | choose the symbols | step | inquiry learning |
| | Floor keyboard | symbols ordering | step combination | challenge |
| | Theme cards | card-map connection | step | inquiry learning |
| | Geometrix | creation of geometric forms | step or object positioning | inquiry learning |
| | Memory | place-content connection | step combination | memory game |
| | Flowers | walk | walk and step | sensory game |
| | Poll | answer | step | information display |
| | Scratch | remove scratch layer | step | sensory game |
| | Football | kick the ball | walk or run ? | real game reproduction |
| | Paper lines | conceptual connections | step combination | challenge |
| | Slides | content management | step | information display |
| | Pong | moving the bat | walk or run | real game reproduction |
| | Rubber ducks | make ducks squeak | step | sensory game |
| | Piano | select a key and see the notation | step | inquiry game |
| | Bubbles | make bubbles burst | step | sensory game |
| | Categorize | conceptual connections | step | challenge |
| SMALLab | Character profile | select personal information | wand head, heart, feet | information display |
| | Chemistry Titration | put elements in the flask | wand drag and drop | inquiry learning |
| | Color Mixer | mixing RGB components | wand head, heart, fee | challenge |
| | Constant Velocity | moving along abstract representation | keyboard or remote controller<br>moving wand | inquiry learning |
| | Constant Acceleration | comparing data with movement | keyboard or remote controller<br>moving wand | inquiry learning |
| | Disease Transmission | experiment various transmission models | remote controller and wand drop and drag | inquiry learning |
| | Fraction Lab | moving along abstract representations | keyboard or remote controller<br>moving hands (kinect) | inquiry learning |
| | Fraction Action | using data in a timed scenario | keyboard or remote controller<br>moving hands (kinect) | inquiry learning |
| | Gear Ratio Explorer | generate gears | distance hand-shoulder | play-acting |
| | Gear Ratio | matching movement with gear ratio | distance hand-shoulder distance | play-acting |
| | Geology Layer Cake | matching objects<br>earthquake generation | wand drag and drop<br>shake handheld devices | inquiry learning |
| | Le Tour de Force | matching movement with strenght | spinning arms | play-acting |
| | Lever Fulcrum Explorer | mechanics concepts manipulation | arms and shoulders | play-acting |
| | Lifting Gears | mechanics concepts manipulation | spinning arms | play-acting |
| | Light and Mirrors | element manipulation | wand drag and drop | inquiry learning |
| | Lightwave | wavelenght-color relationship manipulation | wand drag and drop | inquiry learning |
| | Memory | place-content connection | wand drag and drop | challenge |
| | Oscillating Spring Mass | physics parameters manipulation | wand movement | inquiry learning |
| | Order Line | ordering symbols | ? | challenge |
| | Particle Predator | rule decision | wand drag and drop | challenge |
| | Projectile Flight Game | rule decision | wand movement | challenge |
| | Projectile Flight Graph | strike a virtual ball | wand movement | inquiry learning |
| | Red Rover | catch and crush | ? | challenge |
| | Storyline | image and sound creation | ? | inquiry learning |
| | Sunshine Hearth | orbital position manipulation | wand movement | inquiry learning |
| | Traffic Jack | lift level | arm length | play-acting |
| | Venn Diagrammer | conceptual connections | ? | challenge |
| | Word Play | quiz solution | ? | challenge |
| Stanza Logomotoria | Resonant memory | place-content connection | step | memory game |
| | Fiaba magica | storytelling | raising arms | impaired users |
| AmI Playfield | Apple Hunt | mathematic operations | way-finding | challenge |
| LLP | Newtonian simulation | modelling force and motion | play-acting | play-acting |
| STEP | Water particles | modelling states of matter | play-acting | play-acting |

**Table 3.1:** *Table of responsive floors learning games, grouped by system. Aim, interaction modality and typology are reported for each game.*

# Chapter 4

# A Conceptual Framework for Motion-based Music Applications

In this Chapter a conceptual framework for motion-based music applications is presented. The framework assumes that body movements are the primary interaction modality in interactive spaces. Thus, three key points are determined and discussed: the spatial positioning of interactive landmarks, a reason to choose an interactive landmark target instead of another (i.e., where to move) and a reason to move from one interactive landmark to another (i.e., when to move). These three points form the cornerstones of the analysis of the case studies that will be presented later. To support the importance and the role of bodily movements as primary interaction modality in interactive spaces, an analysis of post-WIMP interfaces, reality-based interaction and related themes is proposed.

## 4.1 Introduction to Post-WIMP Interfaces

Windows, icons, menus, and pointer devices (WIMP) interfaces are the legacy of research carried out since the early seventies at Xerox PARC (California). When made available to the general public by the Macintosh in 1984, WIMP interfaces represented a big step forward in human-computer interaction development. Compared to the previous generation of command-line interfaces, WIMP - or Graphical User Interfaces (GUI) - are easy to use thanks to the employment of icons and pointing devices, like the mouse. They offer to the unexperienced user an easy and simple way to access to computer's content and functions, but they are limited to one-user and one-desktop interaction mode and allow only discrete event input like key presses and mouse selection (van Dam, 1997). While WIMP interfaces are still very popular and commonly used in desktop computers and laptops, a new generation of interfaces has been carried out to help a greater portability and pervasiveness of electronic devices. The so-called post-WIMP interfaces include handheld devices, augmented and virtual reality environments, multi-touch surfaces, and smartphones. This type of interfaces reverse the WIMP paradigm and offer completely new perspectives to the user. In the WIMP interfaces the user enters the window that represents the digital space and acts inside it. Her/his attention is completely absorbed by the machine and the

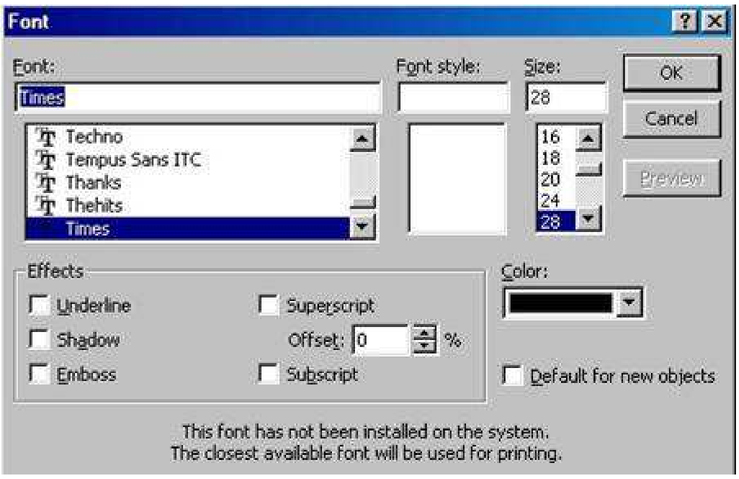



**Figure 4.1:** *A typical graphical interface, with windows, check boxes, and drop down menus. Source* `http://www.teach-ict.com`

environmental influence is confined to the narrow space occupied by the screen, keyboard, and mouse. On the contrary, in post-WIMP interfaces the window protrudes in the user's physical space and is manipulated accordingly. The digital interface becomes embedded in the reality and the way users act in the real world becomes the new paradigm for interaction, including tangible manipulation, spatial interaction, group behaviour and expressive representation of items (Hornecker and Buur, 2006). A good example of this interaction style is provided by Reactable, an electronic music instrument developed since 2003 by the Music Technology Group at the Universitat Pompeu Fabra in Barcelona, Spain (Jordà et al., 2007). In Reactable a translucent surface acts as a collaborative interface where users manipulate custom objects, called tangible, which represent various digital sound synthesis processes. The tangibles offer an expressive graphical representation of the different processes they activate as soon as they are positioned on the instrument's surface. The spatial connections of the various tangibles determine the sound synthesis processes in real time and offer a visualisation of what is going on. This is an example of what Jacob et al. (2008) define as "reality-based interaction", that is the way how users employ the physical world properties in the manipulation of digital content.

### 4.1.1 Reality-based Interaction

The aim of Jacob et al. (2008) was to propose a unifying vision for various interaction styles whether they are virtual, mixed and augmented reality, tangible interaction, ubiquitous and pervasive computing, context-aware computing, handheld, or mobile interaction. However, it must be observed that the dimension of the device involved in the interaction has also some relevance with respect to the type of relationship users can build with the physical world. As an example, handheld devices and smartphones represent reality-based interfaces which are relatively small. They are operated in the user's hand and can be moved all around the environment. This entails

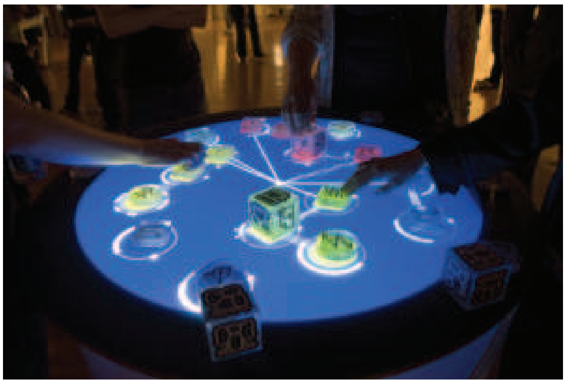


**Figure 4.2:** *Users moving blocks on the Reactable electronic music instrument surface. Source* `https://en.wikipedia.org/wiki/Reactable`

that they are particularly fit for context-aware interaction where an augmented reality interactive space is produced depending on user's location (Makris et al., 2013). This type of interactive space allows also for some social sharing experience, in the sense that co-located users can share the same augmented reality experience at the same time. Another occurrence is when the digital space window has the dimension of interactive white-boards or large touch-screens. In this case the items spatial organisation allows users to interact with the digital content, which lies within an arm-length range and which is reached mainly through fingers' interaction. As many users can assist and take a part in the interaction in the same time, various systems have been experimented to allow documents sharing through various surfaces (e.g., moving a document from a private tablet to a public touch-screen with a simple, intuitive sliding movement) (Wigdor et al., 2006) or for organizing entire interactive rooms where various visualization devices are connected (Jetter et al., 2012). In these environments the user is surrounded by various devices, but s/he is still in a real unresponsive space. When the interaction window further enlarges and becomes a floor, that is a surface where the user can actually enter and upon which s/he can actually walk, the interaction conditions change again. The device is not only surrounding the user, but the user's itself and her/his peripersonal space lay inside the digital window. Moreover, this window can be simply a flat space (like a responsive floor under a camera range) but it can also be a three-dimensional space, where the whole body limbs are employed as a generator of motion data. This means that the reality around the user becomes the source of data that can feed any digital process and that every behavioral model belonging to real life can be used as an interaction modality.

### 4.1.2 Reality-based Interaction Themes in Interactive Spaces

Jacob et al. (2008) identify also some physical world themes which help to understand how reality forges interaction in post-WIMP interfaces. These themes are the following:

- *Naïve physics* that takes in account the effects of real environment physics laws like strength,

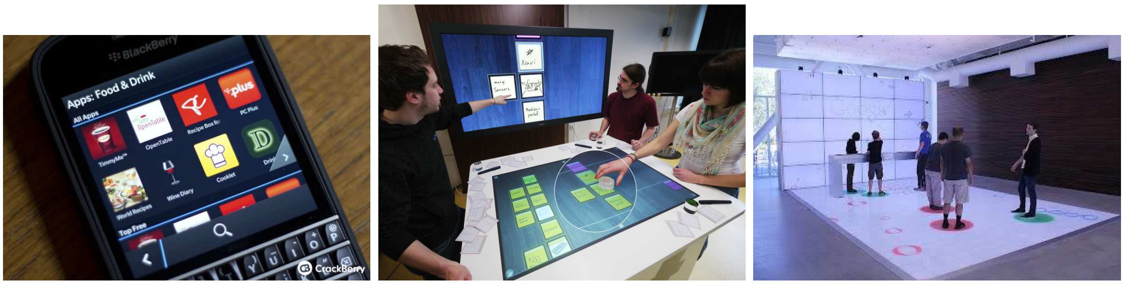



**Figure 4.3:** *Three different sizes for digital windows in interactive spaces. Form the left a context-aware device, a tabletop and a large resolution display of the AffinityTable prototype (Jetter et al., 2014), and the responsive floor of Google's Interactive Spaces (source* `arstechnica.com`*).*

friction, gravity and so on (e.g., scrolling through a menu that simulates the resistance as if it were a spring, or widgets animations with mass and stiffness simulations);

- *Body awareness and skills* that take in account the perception users have of their body in space, e.g., proprioception (the sense of the body and limbs position) and peripersonal space perception;[1]
- *Environment awareness and skills* that relate to spatial cognition that users employ when moving around the world in their everyday life;
- *Social awareness and skills* are the abilities to relate to others and comprehend local co-presence, social affordances and sense of social presence.

These themes are employed to understand the analogy between real-life events and the augmented-reality experience in an interactive space. Particularly, they are reinterpreted in order to provide a basis for the construction of the conceptual framework for motion-based music applications presented below.

#### 4.1.2.1 Naïve physics

The naïve physics examples presented above relate to the digital representation of objects behavior in the natural environment. In an interactive space the user's body is at the origin of the interaction and, consequently, real-world movements of one or more bodies and limbs are the focus. The characteristics of these movements may express data that can be used to produce some augmented reality effect like sounds or interactive projections, as depicted in Figure 4.4. Movement qualities expressed in the Labanotation[2] like effort factors (space, weight, time and flow), can help the understand the expressive potentialities of these data. Some examples can

[1] The peripersonal space is the space around the body and in the reach of the limbs.

[2] Laban Movement Analysis, LMA is a method for describing and categorizing human movement. It was proposed by Rudolf Laban in 1928.

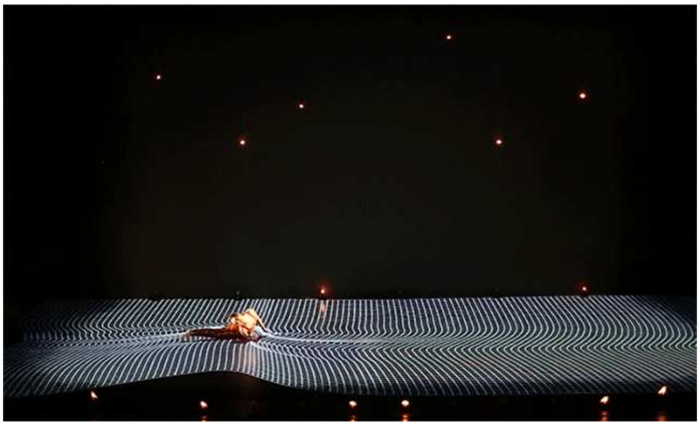


**Figure 4.4:** *An interactive dance setup, with projections depending on dancer's movement. Source* `http://www.thisiscolossal.com`

be found in Mappe per Affetti Erranti, a contemporary dance project based on the already described Eyesweb XMI Library (see Section 2.3.1.2). In this project twelve expressive descriptors (quantity of motion, impulsiveness, vertical and horizontal components of velocity of peripheral upper parts of the body, directness index, etc.) are employed to define four expressive intentions that control the musical output (Camurri et al., 2008). Thus, the physics of full-body movements becomes the main interaction modality in an interactive space and the quality and richness of this interaction is particularly interesting for musical environments where expressivity plays an important role.

#### 4.1.2.2 Body awareness and skills

As the user of an interactive space affects the environment through her/his movements, body awareness and skills play a fundamental role in the interaction. One of the skills users employ in their everyday life is embodied cognition, which expresses aspects of user's experience that are grounded in the physical body and sensorimotor memory. Relevant psychological studies in the field of cognitive and pedagogical sciences, like Fodor (1983), Clark (1998) and Piaget et al. (1952), emphasize the importance of sensorimotor information as part of cognitive processes fed by body experiences in the world. Leman (2008) in his studies emphasizes the role of embodied music cognition as an effective mediator between music knowledge and technology. Thus, not only there is no separation between the abstract and the embodied mind but, as George Lakoff and Mark Johnson affirm, many abstract concepts can be understood through image schemas which are stored in our sensorimotor memory and which come from bodily interactions (Lakoff and Johnson, 2008). Some of these schemas can properly describe some important spatial characteristics of abstract concepts, and, in particular, they can be adapted to depict musical features. For instance the source-path-goal schema expresses the quality of melody proceeding by organizing the experience of goal-directed motion; or the container schema can very well represent the concept of boundary regions like that of tonal harmony (Brower, 2000). More simply, one of the most elementary musical structures, the scale, is commonly described with terms like "ascending" and "descending", which are body-movement and spatial metaphors. In experts as well as in common-people language melodies "go up or down" or "jump" or "arrive" to a particular point. Tones can proceed by "skip" or by "step", keys can be "near" or "far", harmonies make "turnarounds" and bass lines are sometimes defined as "walking basses". All these examples show how musical abstract concepts are deeply grounded in users sensorimotor information and, at the same time, introduce the theme of spatial cognition as linked to embodied cognition.

#### 4.1.2.3 Environment awareness and skills

Environment awareness and skills refer to how a user perceives and manages environmental elements functions and spatial organization. This knowledge is composed of spatial properties like location, size, distance, direction, separation, connection, landmarks and route information, and is acquired via sensorimotor systems that operate as people move about the world. Also the ability to detect environmental acoustic and non-acoustic events is fundamental to coordinate actions according to ecological occurrences (Phillips-Silver et al., 2010).

#### 4.1.2.4 Social awareness and skills

Interactive boards or large touch-screens allow to extend the traditional WIMP interaction to more than one user. When the interactive space is so large to be not only visible to a great number of bystanders but also to be acted by many users, the users actions acquire a completely different meaning. The social sharing of the interactive experience and co-located users participation to the reactive environment exploration makes the consequent cognitive processes a transparent, public event. This may lead to various outcomes as social facilitation,evaluation apprehension, audience enjoyment and emotion arousal enhancement (De Kort and Ijsselsteijn, 2008).

## 4.2 The Design Process of Motion-Based Music Applications

In Section 4.1 post-WIMP interfaces and connected reality-based interaction styles and themes have been analyzed with the aim of understanding the various aspects involved in the interaction. The aim of this Section is to offer an overview of the design process of motion-based music applications. These applications employ bi- or three-dimensional augmented reality interactive spaces, where the user acts and moves in the reality but, through her/his movements, can reach and interact with expressly arranged digital contents. As a consequence, full-body motion can be regarded as the leading factor for human-computer interaction in these environments. The application's design process starts from the *Application's aim* which is connected to the *Musical Information* to be represented (see Figure 4.5). This module can be based on the model of a traditional acoustic instruments, the representation of an abstract musical concept, new spatial arrangements of traditional musical features or completely new virtual instruments at all. In any case, a precise representation of the musical content is needed to extract a valid *Geometric Interpretation*. It is important to underline the word "interpretation" because here lies the core of this

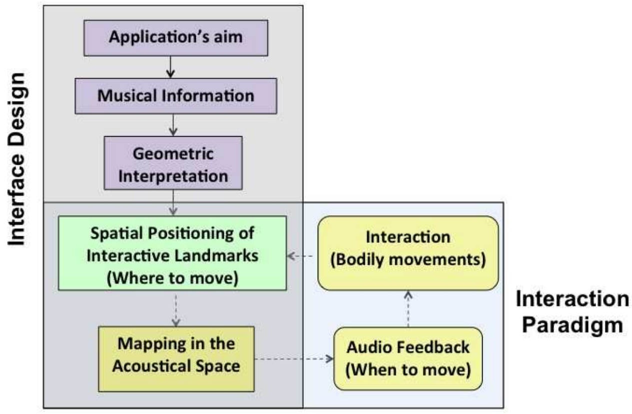



**Figure 4.5:** *From the musical concept to the application's interface. The process starts form the application's aim and from the musical information module and, through a geometrical interpretation, a spatial positioning of some interactive landmarks is arranged on the application's interface. The user moves on it and produces some audio output which influences her/his movements.*

approach to musical applications design. In this phase, the focus is on the musical functionality upon which the application is based and on the main aspect to be outlined. As a example in Hsu et al. (2014) a guitar model is described in a virtual scene, where subsequent blocks positioned on the guitar's neck represent the different chords, whereas another area at the center of the guitar body is used to trigger the sound. The main difference from the real model is that in the guitar the player selects the chords through different fret finger positions, whereas in this digital reproduction a series of simplified chord blocks are proposed. It is clear that a trade-off solution between reality and its virtual environment projection has been found. Also if this solution can fit well from the computational point of view, it is important to point out that in designing such kind of applications always some aspects of the reality must be rendered by interpreting or summarizing their functionalities. Furthermore, as this rendering happens in a virtual environment, a geometrical arrangement of this knowledge has to be produced. This spatial arrangement has to follow precise usability and learnability qualities, necessary to link the theoretical knowledge to its representations. This aspect is particularly important in order to guarantee musical coherence and suitability, mainly when the application aims at social or communication purposes, like in educational environments. Here, the above cited human-computer interaction features must be further tested in order to assess the application's efficiency in conveying an actual musical knowledge to its users. The *Spatial Positioning of Interactive Landmarks* module represents the connection between the *Interface Design* and the *Interaction Paradigm* area. It is the core of the conceptual framework, because here lies the actual application's interface. The positioning can be in the three-dimensional space or on two-dimensional surfaces as in the case of responsive floors. It is is marked by auditory and, optionally, by visual tags or interactive graphical elements projections. The *Mapping in the Acoustical Space* module provides the link between the interactive landmarks and the audio output of the system and, consequently, to the audio feedback. The *Interaction Paradigm* area shows how the system is activated by the user's physical movements onto the interface and how her/his movements are influenced by the audio feedback. The interaction mechanism - involving the direction and the timing of movement - is explained in Sections 4.3.2 and 4.3.3.

## 4.3 The Three Key Points of the Conceptual Framework

The aim of this Section is to offer a comprehensive analysis of the nature and quality of full-body interaction in environments where music is produced or heard. Motion can be described as the kinematic relationship between two points in space. Usually, this relationship involves some physics concept like displacement, distance, speed, acceleration, and so on. When related to the user's change of position in a musical interactive space - which is produced by an artificial, technology-based computer application - motion implies the existence of the following three elements:

a) the spatial positioning of interactive landmarks;

b) a reason to choose an interactive landmark target instead of another (i.e., where to move);

c) a reason to move from one interactive landmark to another (i.e., when to move).

These are the three key point of the conceptual framework that will be analyzed and discussed in the next Sections.

### 4.3.1 Spatial Positioning of Interactive Landmarks

In motion-based applications there is a deep relationship among physical space, user's spatial cognition and digital contents. Thus, some tool to link all these elements together and to manage them in a successful way is needed. An example is provided by the Zone Tracker application (see 2.3.1.2) which employes various masks for the division of the active floor surface (Figure 5.10). Each mask provides a generic spatial organization, with a number of available landmarks where to position the content (e.g., audio files, digital sound processing effects or music composition algorithms). The spatial positioning of interactive landmarks may be visible or invisible to the user. When visible, interactive landmarks may be labelled by visual tags or visualized through graphical elements projections. In any case, the available landmarks must be connected to the content through some spatial organizing principle. This is provided by the "conceptual integration" between the spatial characteristics of musical features and the actual space. The term is borrowed from the blending theory (Fauconnier and Turner, 2008) that is a conceptual framework of human knowledge suggesting that the brain produces information through various forms of integration between two input mental spaces. The theory defines a four-space model: a

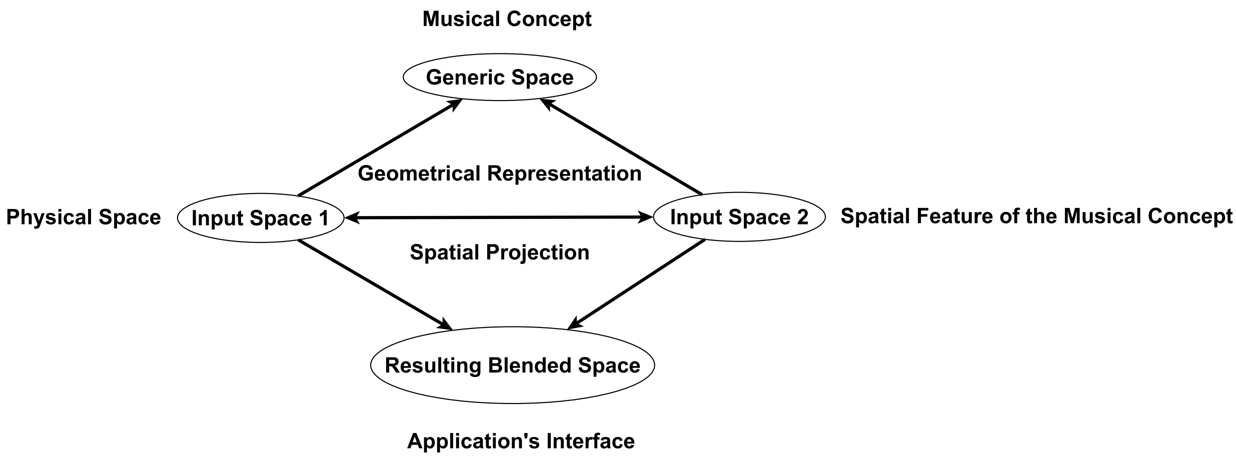



**Figure 4.6:** *Conceptual blending diagram, arranged from Benyon (2012, p. 223). Starting from the generic space of the musical concept a conceptual integration is operated between the physical space and the spatial feature of the musical concept. The resulting blended space is represented on the application's interface.*

generic space, two input spaces and a resulting blended space. The generic space is where some abstract knowledge about the domain is stored; the input space 1 is the physical space where interaction happens; the input space 2 is the musical concepts spatial organization; the resulting blended space results in a spatial projection which takes some characteristics from both input spaces and which creates something completely new (Benyon, 2012). The conceptual blending theory is the basis of many design approaches in responsive floors learning applications, where student's first-person embodied experiences are employed to model physics events in augmented reality environments (Enyedy et al., 2015). As showed in Figure 4.6, in motion-based music applications the physical space has to be matched with the musical concept upon which the application is based. The relationship between the two input spaces is mediated by a geometrical representation of musical concepts and their spatial projection, which provides the sounds (or sound processing zones) positioning in the new blended space. Many musical features like harmony or melodic movements have been historically depicted by spatial representations. Some examples are the Euler's tonnetz (literally "web of tones", a spatial schema showing the triadic relationships upon which tonal harmony is based), or the Gregorian Chant Neumatic Notation,[3] or Chironomy.[4] This suggests the idea that the spatial positioning of sounds conveys some meaning about their inner nature and element organization and that this meaning can be made available for the user through spatial representations.

[3]In the Gregorian chant notation melodic cues are represented by neumes, which are groups of notes tied together to indicate their reciprocal relationships and expressive meaning.

[4]Chironomy is a gestural system expressing melodic contour variations through the conductor's hands motion.

### 4.3.2 Where to Move

To move in space a user needs to know where to go, that is, s/he needs to know where her/his target interactive landmark is. While this information is most of the times obvious in common life where our motion is usually directed to well-known precise goals, the same cannot be said for the choice of a target interactive landmark in an artificial environment. In such a situation motion implies the creation of a cognitive map, which includes landmarks, route segments and wayfinding (Montello, 2001). But a cognitive map is much more than a simple mental routing sketch, as it includes other non-spatial elements, like perceptual attributes and emotions aroused by the system's audio feedback. Indeed, the creation of the new blended space allows the user to navigate musical concepts in the physical space and to move literally inside them, obtaining a different audio feedback depending on the occupied zone. The coupling between concepts spatial location (interactive landmarks) and musical-content perception begins to feed the user's cognitive map, which will drive her/his decisions about where to move in the artificial environment. This coupling is also driven by the musical language implicit knowledge, which is a well-known mental process that allows the unconscious acquisition of very complex and structured language constructions. The mind is continuously fed by structured stimuli (language, music, spatial relationships, sensorimotor information, etc.), and, independently from user's will, builds an inner knowledge about them (Reber, 1989). As an example, the public of the World Science Festival 2009 employ the implicit knowledge of the pentatonic scale to couple Bobby McFerrin's movements to the sounds.[5] On the other hand, considerable research in the field of music psychology has offered evidence that even children from the early age of 4–5 years have a wide implicit knowledge of tonal harmony too, comprehending chord functions, harmonic relations, and perception of regularities of harmonic frames in time (Corrigall and Trainor, 2010). As tonal music is ubiquitous in our music culture's environment, all these features are learned from mere passive exposition and can so be considered as a cognitive heritage common to every user, independently from her/his degree of musical education (Tillmann et al., 2000). Thus, when a user enters a music interactive space, her/his implicit musical knowledge is elicited by the system audio output and can so be accessed, used and/or modified by the user during the experience.

### 4.3.3 When to Move

The link between music and motion is called entrainment, which is an organism's ability to synchronize movements when in the presence of an external rhythmic stimulus. Entrainment is the process responsible of "when" to move in the interactive space. In physics, entrainment is defined as the frequency alignment of two oscillating bodies on phase or at 180° out of phase. It happens because small quantities of energy transfer from one body to the other until they are fully synchronized. In human beings, the firing frequency of auditory neurons when receiving an external rhythmic stimulus influences the firing frequency of motor neurons, causing in this way coordinated movement (Thaut et al., 2014). This resonates with the model of entrainment

[5] Bobby McFerrin is a famous american jazz vocalist and conductor. The reported example refers to a performance at the World Science Festival to show the role of musical expectation (Notes & Neurons: In Search of the Common Chorus). The video can be found at `https://youtu.be/DBJ7mBxi8LM`

proposed by Phillips-Silver et al. (2010), which is composed of three phases:

a) the rhythmic detection of environmental signals (not only acoustical but also as a byproduct of ecological phenomena);

b) the ability of producing rhythmic signals (not only deriving from musical activities but also from other biological activities);

c) the ability to integrate both preceding phases in order to adjust the output's rhythm.

This framework is interesting because it does not limit the idea of entrainment only to the basic regular pulse synchronization generally observed when people clap their hands to their favorite song's beat in public concerts, when music players align their individual timing on the conductor's gesture or when dancers perform the same movement in a strictly rhythmic fashion. It extends the entrainment range towards a more general and wider number of biological and ecological entrainment phenomena like environmental signals (season's or day/night turnover, weather changes, wind blowing, sea waves shattering) or biologically produced rhythms (breathing, eating, heart beating, walking, crickets' chorusing, wolves' howling and so on). Many living beings' actions depend on various form of entrainment with these signals, both in natural and in artificial environments (Erkut et al., 2013). These examples emphasize at least three important aspects of entrainment:

a) entrainment may occur in conditions also very different from regular rhythmic input (predictive entrainment);

b) entrainment is connected not only to "when" to move, but also to "why" to move (selective entrainment);

c) coordination is the condition of success in the activity.

#### 4.3.3.1 Predictive entrainment

To understand how entrainment works in a musical interactive space it is necessary to focus not so much on rhythmic input regularities, which can be found both in natural and in artificial environments, but rather on rhythmic predictability. A regular pulse has a high degree of predictability. Nonetheless, also a pulse slowing down (rallentando) has a less but effective degree of predictability because the listener can follow a previously stored model of rallentando and try to adapt it to the actual event sequence (Friberg and Sundberg, 1999). Another example of how humans can adapt entrainment to particular situations is the case of "soft entrainment", which occurs when little deviations in rhythmic entrainment among different performers in music ensembles are registered. Soft entrainment may occur at various degrees of deviation, depending on the phase of the musical phrase. Yoshida et al. (2002) report that synchronization is maximal when aiming at the phrase's climax (tense phase), whereas it deviates more and more while approaching the phrase's end (relaxing phase). These examples show how entrainment is a dynamic process, which involves not only mere beat detection but also a much wider range of musical elements, like expressive trends, motion patterns and musical phrase organization. Jones and Boltz (1989) provide an extensive framework of how real world time structures are organized in a hierarchical way, allowing so predictive entrainment to work. They affirm that the distribution of many natural events' markers is nested at various time levels that are consistent with ratio or additive time transformations. This explains not only why humans can entrain with natural phenomena like gradual or abrupt changes of velocity, but also how prediction works when they have to synchronize with multi-level, hierarchically-organized time events. Musical metric structure for example starts from a lower level composed by the smallest rhythmic units and, through successive layers stratification, reaches much extended musical units like musical periods, forms, sonatas, or symphonies. This hierarchical organization allows a subject to have an idea on how musical events are organized and to make a prediction about the time s/he has to wait until the appointed event. Yet, a wider look on musical entrainment has to include as well the observation that not all kinds of music are strictly based on isochronous pulse, just as not all the parts of a beat-based music are rigorously dependent on beat. Think of a classic concert's cadenza, where the soloist leaves the overall ensemble governing pulse playing freely to express all her/his virtuosistic ability; or of the Gregorian chant's swinging gait, where a pulse can sometimes be perceived, but always among many breathing pauses and fermatas.[6] There are also types of beat-composed music where the pulse is not perceivable at all from the musical output. This is the case of many classical contemporary music compositions (like Ligeti's Lux Aeterna, just to cite one of the most popular works of this genre), where the lack of pattern periodic repetitions prevents any musical elements metrical organization. Notwithstanding, all the cited examples show a more or less high degree of predictability, because, also if the events are not subjected to a regular metrical organization, they alike show some shared musical or non-musical pattern. In classical music solo and orchestras concerts, the solo's cadenza is marked by a precise fermata on the second inversion of the I degree at the beginning and by a conclusive dominant chord in root position at the end. Thus, two harmonic markers act as strongholds of the relatively beat-free event, allowing the conductor and the whole orchestra to re-synchronize their beat at the end of the cadenza. In Gregorian chants the predictive timing of events is given by the musical phrases breathing times, whose code is deeply grounded in physiological, expressive, and melodic structure cues. Chaotic mass movement produced by very small musical elements are the result of physics model-based composition techniques employed by many $20^th$ century composers. One example is Ligeti's micropoliphony (Bernard, 1994), where the musical elements are controlled by tendency masks that rule events' rhythm, density, and height. The temporal evolution of a tendency mask is perceived by the listener in ecological and physics terms like growth, proliferation, thickness, fluctuation, and so on. Thus, process end points are perceptual landmarks that can be regarded as predictable entrainment anchors.

#### 4.3.3.2 Selective entrainment

Selective entrainment is the subject's ability to focus her/his synchronization activity on a specific environmental signal, chosen among multiple simultaneous rhythmic stimuli. It is easy to notice

[6] The fermata is a sustained note, chord, or rest, whose duration is longer that the indicated value.

that in natural environments the acoustic or non-acoustical signals overlapping is the common case. For example it is not difficult to imagine that a hunter has to select her/his prey's biological signals among all the other signal in the surroundings to be successful in the chase. Hence, the subject has to filter among the incoming signals and to focus on a specific input depending on the goal s/he wants to achieve. This means that movement is triggered by some environmental changes which the subject is interested to, and that movement depends on the perceptual timing of these changes. Thus, "when" to move is strictly connected to "why" to move, whereas the deep reason to take off is the need to remain tuned to dynamic, basic events. Whereas following ecological signals may have very important biological consequences in the natural environment, it acquires all a different meaning in artificial interactive spaces where signal flow is controlled and where user's responsiveness is one of the fundamental parts of the application's design. First of all, if the interaction logic is controlled by the designer, the reaction to the signals is always mediated by the user's implicit knowledge, about which the designer can make no more than common sense guesses; secondly, in case of multi-level signals like musical input, the user is called to apply a selective entrainment as s/he has to decide at what level to synchronize with the input. The general idea about entrainment in artificial environments is that movement is always the result of a cognitive selective process related to previously acquired knowledge. The consciousness about where to direct attention is the key element of selective entrainment: How can designers help users in achieving this goal, remains a great challenge in reality-based interaction design.

#### 4.3.3.3 Cognitive meanings of coordinated movement

No entrainment activity in natural or artificial environments would be successful without coordinated movement. Whether it is a tight or softer coordination, it is clear that action success depend on the subject's ability in detecting the right rhythmic input and to align her/his behavior to it. Baimel et al. (2015) underpin how behavioral synchrony in collective activities fosters social coordination and empathic concern, stressing how tuning minds together helps individuals in achieving abilities. Moreover, Clayton et al. (2005) observe that cognitive activities like perception, attention and expectation depend all on entrainment, and that, particularly, musical entrainment helps motor and self-control in individuals. These observations foster an idea of entrainment as a general cognitive key that rules human relationships with world's events. Considering entrainment in an interactive space, the above cited framework from Phillips-Silver et al. (2010) could be re-interpreted in the following way:

a) the rhythmic detection of environmental signals represents the openness of the subject to receive information from the system output and to select where to direct her/his attention;

b) the ability of producing rhythmic signals means that the user understands how the application's interaction mechanisms works;

c) the ability to integrate both preceding phases in order to adjust the output's rhythm establishes the point where the system's response may influence the user's behavior, making her/him learn something new.

# Chapter 5

# Responsive Floors Music Applications

In this Chapter, three case studies of responsive floors music applications are presented. Harmonic Walk is a music learning application devoted to the study and practice of western tonal harmony. It belongs to the category of play-acting games, because while performing the application's ultimate task - the melody harmonization - the user moves in the same way and in the same time as the musical chords do in the harmonic space. "Good or Bad?" is a challenge game where two users cooperate in recomposing a reference musical piece by selecting or discarding some musical tracks belonging to the reference piece itself or to an antagonist piece. Users employ their musical listening skills to complete the game task with the fewest number of errors. The third application - Jazz Improvisation - is an active listening application belonging to the category of inquiry games. A number of freely stackable music tracks are laid on various interactive landmarks arranged on the responsive floor's surface. The tracks are activated by the user's passage and are muted in the second step.

## 5.1 Case Study 1: Harmonic Walk, a Music Learning Application

Harmonic Walk is a responsive floor environment designed for learning and practicing the accompaniment of a tonal melody. Employing a video camera motion tracking system, the application offers to the user the possibility of getting in touch with some fundamental tonal music features in a very simple and readily available way. Notwithstanding tonal music is very common in everyday life, musically untrained people as well as music students and even professionals are scarcely conscious of what these features actually are. Harmonic Walk through bodily movements in space can provide all these users with a live experience of tonal melody structure, chords progressions, melody accompaniment and improvisation. Moreover, it can return to users a physical feedback of the spatial relationships which govern the music chords domain, thus helping them to build their personal cognitive map of tonal harmony. Harmonic Walk, matched with interactive floor graphics, can be used as a didactic tool to support music education programs.[1]

[1] An example of Harmonic Walk utilization can be found at `https://youtu.be/c4ru468eqM0`

**Figure 5.1:** *Harmonic Walk while being tested by three high school students at the Catholic Institute Barbarigo, Padova (Italy).*

### 5.1.1 How it Works

The Harmonic Walk's aim is to drive the user towards tonal melody harmonization, which is a complex and multi-faceted task for the user. The ultimate goal of melody harmonization may be preceded by preparatory activities like the discovery of the melody harmonic rhythm and the exploration of the harmonic space. The melody harmonic rhythm is the duration of the harmonic regions which correspond to the various chords employed in a song. The song audio file is cut in correspondence of the harmonic changes and the resulting audio fragments are laid one after the other along the active surface's borders. The user is asked to link the harmonic regions by stepping to the next position in time with the harmonic change. If s/he moves early, the audio fragments overlap; if s/he moves late, the song is interrupted. In the exploration of the harmonic space, the user is asked to move freely on the application's surface where six chord representing a major tonality harmonic space have been arranged. S/he has to match her/his position to the sound of the various chords, to create a cognitive map of the environment which can drive her/him to accomplish the melody harmonization task. At the end, while the melody is sung by a teacher or by the students group, the user harmonizes it by moving on the right chord in time with the harmonic rhythm.

### 5.1.2 Related Work

The idea of interacting with harmonies through spatial representations is not new. It has been widely developed in the Harmony Space project at the Music Computing Lab of the University of Stanford (Holland, 1994). The Harmony Space interface shows a desktop bi-dimensional matrix of pitches ordered by major thirds on the horizontal axis and by minor thirds on the vertical axis. Choosing a key area, a chord size and a chord mapping, when a note is selected, the chord built on it will sound. The interface has been used to simplify the study of harmony, to analyze musical pieces and to compose new ones. The environment is very rich and complex as it allows many option possibilities (the number of pitches used for the chord, the chord type, the key, etc.): so the interaction actually depends on a preselected series of options, which rather fits an expert user level. More recent systems like Isochords (Bergstrom et al., 2007) or Mapping Tonal Harmony (mDecks Music, 2012), are aimed at understanding the harmonic structure through a visualization of the space of musical chords or highlighting preselected harmonization possibilities. Again, these are very complex environments which improve musical structure consciousness at a high degree of knowledge. All the interfaces described so far employ some spatial schemata of the harmonic space, but these are always reproduced on the computer traditional bi-dimensional screen. Nevertheless, Holland himself in 2009 tried a physical space extension of his Harmony Space employing a floor projection and a camera tracking system (Holland et al., 2009). A GPS system to navigate a wide tonnetz area has also been proposed by Behringer and Elliott (Behringer and Elliott, 2009), where the authors suggest some musical games about composition and harmony features knowledge. The PaperTonnetz (Bigo et al., 2012) presents a more intuitive approach, where chord progressions are obtained through a pencil navigation on an interactive paper plotted tonnetz. A more recent experience of interaction with Holland's Harmony Space has also been experimented using a three-dimensional graphical representation of the tonnetz. The user interacts with the interface through colored controllers which are used to select the pitches and to produce the chordal audio feedback (Hedges and McPherson, 2013). Also in the Harmony Navigator (Manaris et al., 2013) the chord selection, supported by a corpus-based statistical model, is operated by hand gestures in the three-dimensional space around the user. All these experiences show that the role of the physical space in the tonnetz navigation is not irrelevant. "... deeper engagement and directness, rich physical cues for memory and reflection, embodied engagement with rhythmic time constraints, hands which are free for other simultaneous activities (such as playing a traditional instrument) and qualitatively new possibilities for collaborative use ..." are the benefits of physical interaction outlined by (Holland et al., 2009, p. 7). Others, like Bigo et al. (2012) and Manaris et al. (2013), emphasize the role of gestures as the preferred interaction style in chord navigation, because they can help the engagement of non academic musicians or amateurs with the world of harmony (Hedges and McPherson, 2013).

### 5.1.3 Harmonic Walk Music Information

Harmonic Walk is based on two tonal music's features: the melodic segmentation and the harmonic space of a one-key tonal melody. As soon as a listener is presented to a tonal melody, s/he first tries to interpret the sequence of notes, grouping them after a metrical and harmonic frame (Povel and Jansen, 2002). This produces a segmentation of the composition into different harmonic regions which, in case of one-key melody, belong all to the same tonality. The example in Figure 5.2 shows an excerpt of a popular melody written in the tonality of F major. The gray units correspond to the tonic regions (T), while the violet ones to the dominant regions (D). In the lower part of the Figure, a one line staff shows the durations of the units, which represent the harmonic rhythm of the composition. As can be seen, it is not regular even in such a simple piece of music. The perceptual, cognitive procedures which lie under this musical representation have been studied and formalized by Lerdahl and Jackendoff (1985), who also showed how a tonal piece perception is organized at various levels, from the most superficial (grouping structure) to the deepest (prolongation reduction). Some of these level are shown in the upper part of Figure 5.2, where, from the $1^{st}$ level to the $2^{nd}$level, a time span reduction is operated, based on the fact that the first dominant unit arrives on the weak part of the first bar. So, it can be incorporated at the upper level, creating a larger tonic unit which occupies two bars. This fosters the idea of the great importance of metrical accents, which rule the changes of harmony in a recursive way.

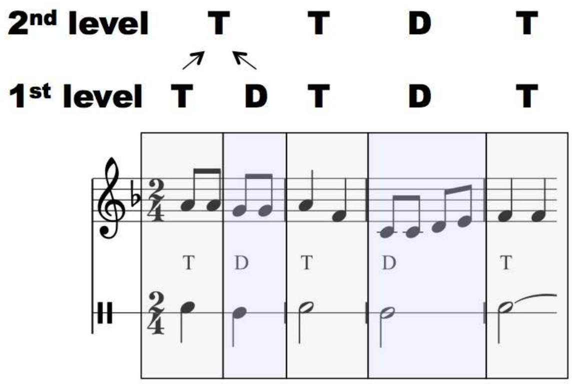



**Figure 5.2:** *A popular melody's harmonic rhythm and levels of grouping structure.*

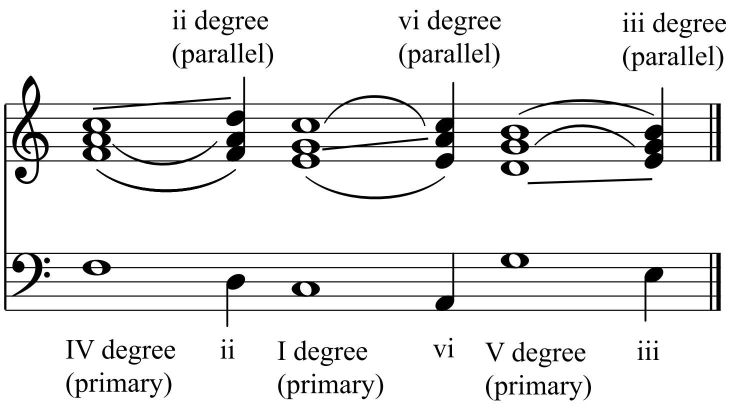



**Figure 5.3:** *Table of the six primary and parallel chords in the key of C major used in the Harmonic Walk application. The common notes between a primary and its parallel chord are tied, while the changing notes are linked with a straight line.*

Summarizing, a tonal piece may be seen as a sequence of musical units, segmented on the basis of their perceived underlying harmonic functions. The sequence has a metrical timing organization, which means that the region's length is always a multiple of a basic time unit. The tonal language is a highly hierarchical system based on the prevalence of the tonic chord matched by the dominant (V degree, one fifth above the tonic) and the subdominant (IV degree, one fifth below the tonic). These three chords are called primary because they represent the fundamental tonal functions in a given key (Bharucha and Krumhansl, 1983). Although every one-key melody can be accompanied using only primary chords, also parallel chords, built one third below primary chords, can be used as supportive harmonies. Hence, the complete table of musical chords used by the Harmonic Walk application consists of six chords, corresponding to the first six degrees of the key, as depicted in Figure 5.3.[2] The tonal harmonic space has been historically represented

[2]The vii degree chord is not included in the Harmonic Walk's chord space because, as a dissonant chord, it requires a particular voice leading. Actually, it is very seldom used in root position and its harmonic function can be easily substituted by the dominant chord.

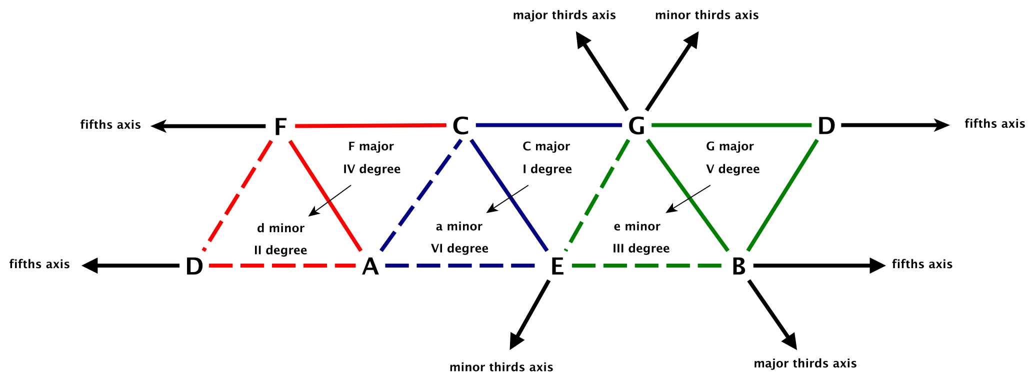



**Figure 5.4:** *The Riemannian tonnetz representation of the three primary and parallel chords (linked with the thin black arrows) in the C major key. On the horizontal lines ($x$ axis) lie the circles of fifths, while on the two diagonal axes the circle of major and minor thirds.*

by the tonnetz. In this representation, the pitches are arranged in horizontal and diagonal lines. As shown in Figure 5.4, the horizontal lines connect the circle of fifths. On the diagonal axis SE-NW lays the major third circle while the diagonal axis SW-NE hosts the minor third circle. The chord is represented by the triangular area resulting from the crossing of the three axes. The tonnetz representation shows the strict connections among the chords belonging to the same key. The whole chord progression of the six C major chords from the left to the right (d minor-F major-a minor-C major-e minor-G major) is originated by the so-called edge transformation. The edge transformation is an inversion of the triangle along one of its edges, which means that two chords sharing the same edge have two pitches in common. Particularly, this relationship links each primary chord with its parallel (with the same color in Figure 5.4), while primary chords (F major-C major-G major) are originated by a vertex transformation, which expresses the commonality of only one pitch. All these observations can explain why the tonnetz representation of the harmonic relationships is so popular and meaningful also from the perceptual point of view.

### 5.1.4 The Geometric Interpretation of the Harmonic Regions Sequence

The time proceeding of the various musical units is led by the melody, whose metaphoric scheme is expressed by the so-called “source-path-goal” schema (Lakoff and Johnson, 2008). Thus, the description of a melody could include a starting point (the source), a series of intermediate steps (all the subsequent musical units) and the end of the musical phrase, usually a cadence to the tonic or to the dominant (goal). Following this metaphor and imagining the simplest motion in space a human can do - the walk -, the tonal composition can be represented as in Figure 5.5, where each step corresponds to the next musical unit. Hence, the geometrical interpretation of the “source-path-goal” schema is a sequence of units arranged along a straight line after which the user can perform the unit sequence till the cadence. To rebuild the piece’s correct time sequence,

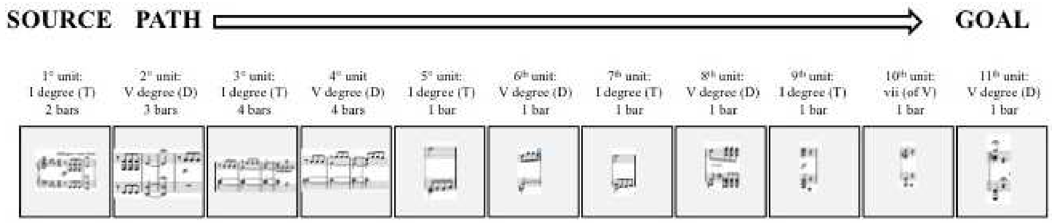



**Figure 5.5:** *The geometric interpretation of the first twenty-one bars of Beethoven's Symphony No. 5 segmented in eleven units corresponding to the harmonic changes.*

her/his pace has to follow the harmonic rhythm of the composition.

### 5.1.5 The Geometric Interpretation of the Six Roots Harmonic Space

As far as concerns the second musical feature, a geometrical interpretation of the harmonic space already defined by the tonnetz along the red, blue and green axes of Figure 5.4 is needed. Table 5.1 shows that all the chords - excluding the problematic vii degree - can be linked with more or less probability with all the others.

Table of Usual Root Progression (W. Piston)

| from | often to | sometimes to | less often to |
|---|---|---|---|
| **I** | *IV or V* | *vi* | *ii or iii* |
| **ii** | *V* | *IV or vi* | *I or iii* |
| **iii** | *vi* | *IV* | *I, ii or V* |
| **IV** | *V* | *I or ii* | *iii or vi* |
| **V** | *I* | *IV or vi* | *ii or iii* |
| **vi** | *ii or V* | *iii or IV* | *I* |
| **vii**$^{6}$ | *I or iii* | | |

**Table 5.1:** *The table summarizes the most used root progression in tonal harmony. Note that the vii degree chord is not represented in root position, but in its first inversion, as it appears most of the times.*

This means that in spite of all its expressive power, the tonnetz is not the best spatial schema to be transferred to the applications's actual surface, because it doesn't allow to move from one chord to another without touching any other chord. Instead, a user must be able to connect all the chords to realize all the possible harmonic progressions. For this reason the geometrical form of the six available musical chords has been reinterpreted, expanding the original tonnetz grid to obtain an inner empty zone, which can allow the transition to the other remaining five chords. The proposed geometrical representation, depicted in Figure 5.6, is a circular ring sliced in six parts corresponding to the six chords. To clarify the relationship between this geometrical form

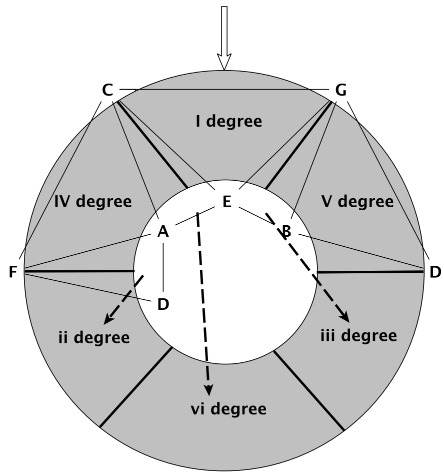



**Figure 5.6:** *The stretched tonnetz with the parallel harmonies overturned in the lower part of the circular ring. The white arrow indicates the position of the Tonic, which is in most cases the beginning of a tonal composition.*

and the tonnetz, the tonnetz chord displacement has been superimposed on the upper part of the ring. The bigger thin lined triangles are the primary chords, while the squeezed ones are the parallel chords. To obtain the required linking space, the parallel chords have been projected in the lower part of the ring. The ii degree is overturned along the F-D axis and is placed just under the IV (following the dashed arrows). The iii degree is overturned along the E-B axis and is placed just under the V degree. The vi degree is overturned along the A-E axis and is placed in front of the I degree. This new displacement provides a logical chord arrangement, where from every chord it is possible to reach each of the remaining five chords without touching the others. Summarizing, the geometrical interpretation of the outstanding tonal melodies features is composed by a straight line and a circular ring.

### 5.1.6 Harmonic Walk Interface

The Harmonic Walk user interface consists of a rectangular area whose dimension change depending on the ceiling height. To help the user's perception of the active area, the straight line of the musical unit sequence is arranged along the borders, while the center of the circular ring is arranged at the center of the rectangle.[3] The user's paths are identified through visual tags which are positioned at the center of each zone. There are two employed paths: the first is the one corresponding to the straight line and is marked with the white crosses, while the second is the circular one, marked with the black crosses (see Figure 5.7). The beginning of each path is marked with an arrow. In this way the user has a visual cue of the center of the various regions and is guided through them by his inner feeling of the music structure.

[3]The masks are depicted as a square and circular ring also if, due to camera pixel shape, the actual masks are respectively distorted in a rectangle and in an oval ring.

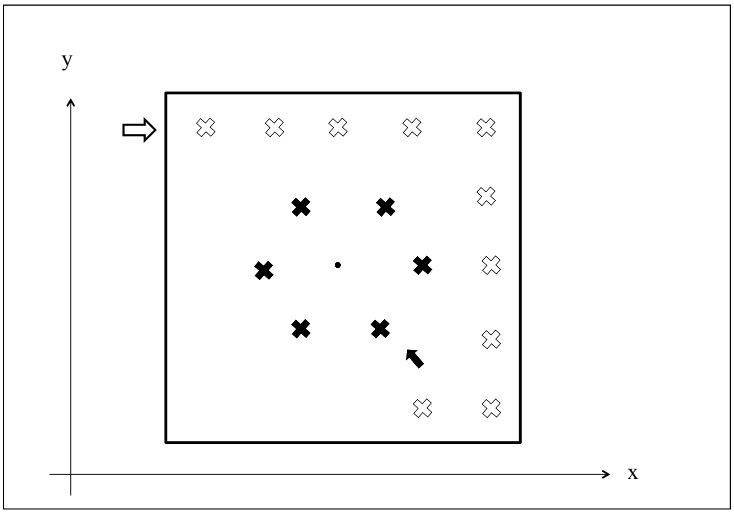



**Figure 5.7:** *Visual tags of the straight and circular path of Harmonic Walk. The straight path (white crosses) follows the interface perimeter and displays a eleven units musical sequence, while the circular path (black crosses), centered with respect to the application's surface, displays the six harmonic space chords distribution. The $x$ axis corresponds to the interface width, the $y$ axis is the interface depth.*

### 5.1.7 Harmonic Walk System Architecture

The Harmonic Walk architecture is composed by two software modules, aimed at video analysis and sound production respectively. The video analysis algorithms and masks for surface division are provided by the Zone Tracker application (see Section 2.3.1.3). The sound production module is provided by a Max/MSP[4] patch. The two software modules communicate through the OSC protocol.[5] A ceiling mounted video camera, oriented perpendicularly to the floor, captures the users' movements inside a rectangular area, whose dimensions depend both from the distance camera-floor and the field of view of the lens (see Figure 5.8). As the system is designed for carrying out educational activities inside a classroom, the camera lens is chosen in order to view a rectangle of about 3x4 meters when the camera is mounted on a 2.8/3 meters high ceiling. The video application Zone Tracker determines the user's position employing a series of masks. These subdivide the active surface, allowing to detect the portion of space occupied by the user. In this implementation of Harmonic Walk, two different masks corresponding to the two geometric representations discussed in Sections 5.1.4 and 5.1.5 are used (see Figure 5.10). The first mask subdivides the tracked zone in twenty-five squares, whose dimensions, after empirical ob-

[4]Max/MSP (`https://cycling74.com/`), is a visual programming language for audio and video production, algorithmic composition and signal processing, written by Miller Puckette in 1980.

[5]*OSC*, acronym of Open Sound Control (Wright, 1997), is a communication protocol based on modern network technology (`http://opensoundcontrol.org/introduction-osc`) was developed, and continues to be a subject of ongoing research at UC Berkeley Center for New Music and Audio Technology (CNMAT). Originally intended for sharing music performance data, it is commonly used to exchange information between computers, applications and various multimedia devices.

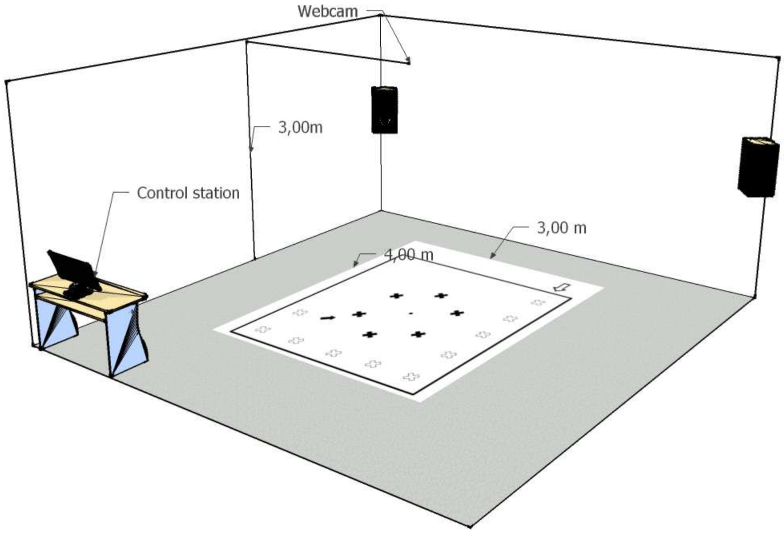



**Figure 5.8:** *The Harmonic Walk's physical environment with the tagged active floor, camera placement, audio monitors and control station.*

servations, has been found to correspond approximately to the distance of a human step. This mask is employed to project the straight path of the musical unit sequence. The circular ring mask is subdivided into six zones. The radius of the outer and inner circles are defined accordingly to the same principles as the dimension of the squares. This mask is employed to project the circle of the six chords of the harmonic space. Zone Tracker sends via OSC the numbers corresponding to the various regions to the Max/MSP patch. Here, for each tonal composition two kinds of audio files are stored: one is the music audio file segmented in correspondence of the harmonic changes units; the second is a group of audio files which reproduce the chords which belong to the song's key chord space and among which there are the chords employed in the song. The chords are played using a MIDI synthesizer with the same rhythm and timbre of the original song employed in the activity.

### 5.1.8 Harmonic Walk Application Framework

Table 5.2 lists all the data that describe the Harmonic Walk application with respect to the conceptual framework reported in Sections 4.2 and 4.3. The musical information, geometric interpretation and spatial positioning have already been discussed in the previous Sections. In the Harmonic Walk application the interaction mechanism is very important because all the knowledge the user acquires is expressed by her/his movements direction and timing. In the first application's activity the user has to follow the straight path which represents the subsequent harmonic regions. S/he has not to worry of where to go, because the regions are put one after the other, but rather has to focus only to when to move. To obtain this information the user has to store in her/his memory the song chosen for the activity and apply on these data her/his previously acquired implicit knowledge of the harmonic changes. As can be seen in Figure 5.11, a

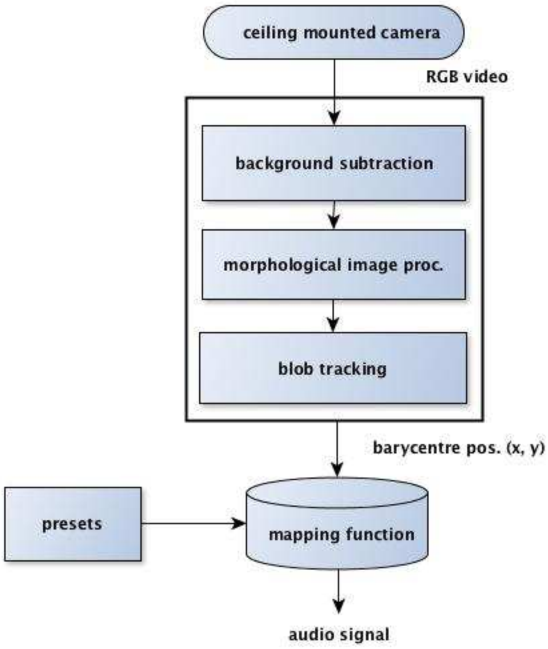



**Figure 5.9:** *Harmonic Walk's system architecture. The RGB video data are processed by the three Zone Tracker modules (background subtraction, morphological image processing and blob tracking). The barycentre position is compared to the stored masks and, through the application's musical presets the audio output is produced.*

musical period offers to the listener various levels at which s/he can synchronize her/his movements. Whereas the period, phrases, semi-phrases and bars levels have a regular timing, the harmonic regions have not always the same duration. Moreover the harmonic region durations overlap the phrase limit in the middle of the period. Thus, the choice to coordinate the step to the harmonic rhythm level (selective entrainment) is not trivial and has a great importance from the cognitive point of view, because it means that the user is aware of the musical element that is the object of the activity, and that s/he has recognized its features. As the songs have words to mark their rhythmic events, some syllables occur also in the harmonic change points. Thus, for reasons of practicality, the syllable which mark the harmonic changes are called the "syllables of change". The syllables of changes are very important to help the harmonic change points identification, memorization and assessment. The exploration activity is characterized by freedom in the movements inside the six zones of the circular ring corresponding to the musical chords harmonic space and by freedom in the timing of this exploration. In this phase the user has the possibility to listen carefully to the sound of the chords and to link this perceptual experience to the various zones of the circular ring. Thus, a cognitive map of the harmonic space useful for the melody harmonization is built. The user is free of space and time constraints when exploring the environment. The observation of her/his behaviors are an interesting source of information about the user's belief and feeling of the harmonic space. The ultimate melody harmonization activity puts together the time constraints of the harmonic rhythm detection and the chord localization

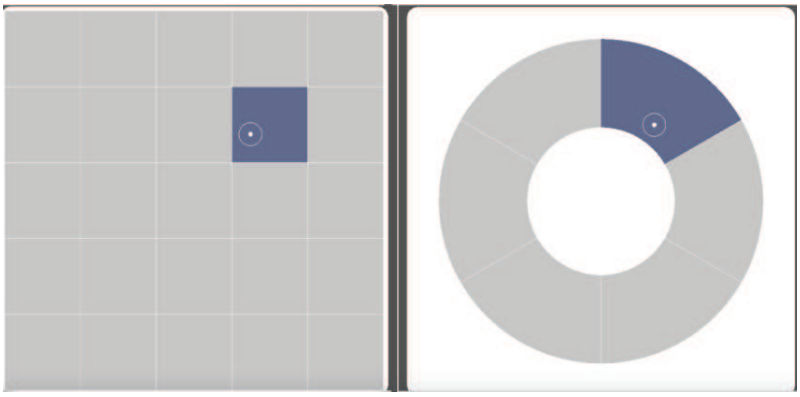


**Figure 5.10:** *Two masks for floor surface division in the Zone Tracker application. A 25-squares zone partitioning and a circular ring with six active zones, designed for button interaction. Blue areas are activated by a user's presence.*

| **HARMONIC WALK APPLICATION FRAMEWORK** | | | | |
|---|---|---|---|---|
| **Aims** | | Discovery of the melody harmonic rhythm | Exploration of the harmonic space | Melody harmonization |
| **Interface** | *Musical Information* | Melody segmentation | Tonnetz representation of a major tonality harmonic space | |
| | *Geometric Interpretation* | Sequence of melodic units along a straight line | Circular ring with three main and three parallel roots | |
| | *Spatial Positioning* | White crosses along the surface borders | Six black crosses put in circle at the centre of the floor | |
| **Interaction** | *Where to move* | Along the straight line | Free choice inside the circular ring | Follow the chord sequence which fits the melody |
| | *When to move* | Selective entrainment with the harmonic rhythm | Free timing | Selective entrainment with the harmonic rhythm |

**Table 5.2:** *The table shows Harmonic Walk aims, musical information, geometrical interpretation and spatial positioning of interactive landmarks employed for each activity. The double vertical bar divides the discovery of the melody harmonic rhythm activity because it employs a different musical information and geometric interpretation than the other two activities.*

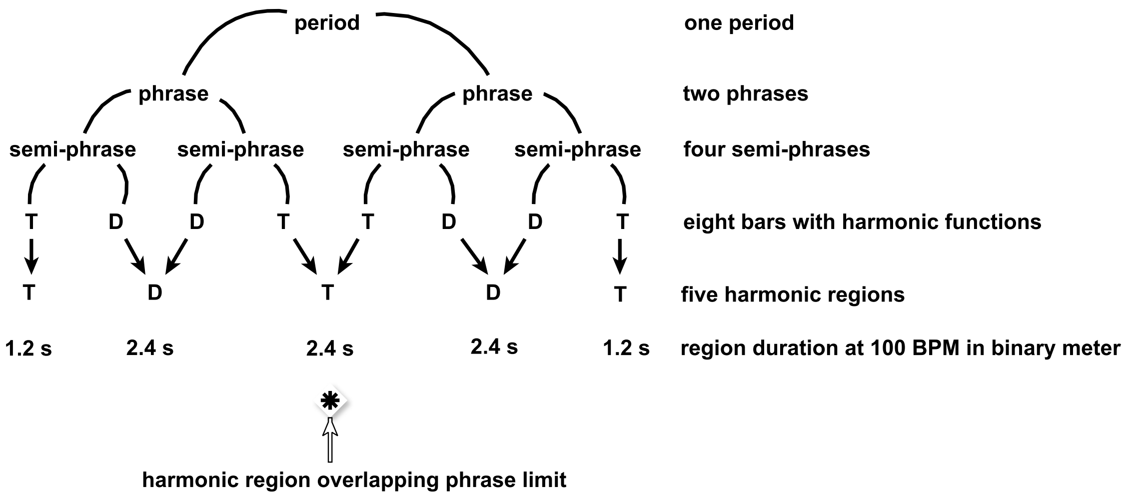



**Figure 5.11:** *Formal organization of a typical music period with phrases, semi-phrases, bars and harmonic functions and regions rhythm. Between the second and the third semi-phrase an overlapping harmonic region is pinpointed.*

stored in the cognitive map built during the exploration activity. In this activity the embodied knowledge of the harmonic framework of the melody is expressed through the sequence of the positions corresponding to the musical chords and through the timing of these movements. The user play-acts the harmonic sequence and behaves as if s/he actually "is" the harmonic sequence. Thus Harmonic Walk may be considered as belonging to the category of play-acting games discussed in Section 3.2.3.

## 5.2 Case Study 2: "Good or Bad?", an Active Listening Game.

"Good or Bad?" (GoB from now on) is a two players game devoted to musical listening and music layer's discrimination. A non-monophonic piece of music is formed by the combination of different musical parts that perform various functions (melody, counter-melody, accompaniment, percussion, bass, ostinato, pedal, etc). These functions can be performed by various instruments (strings, brasses, guitars, keyboards, etc) and can be regarded as a series of superimposed musical layers. For this game two groups of musical layers are employed. The musical layers belong to two different compositions of the same genre (classical, pop, etc.). The two groups, represented with stars and squares, are projected in eight positions on the application's responsive floor reserved to the first player. At each step, a new musical layer is played, and, each time, the second player must decide if the actual layer is compatible or not with the previous ones, by stepping on the good or bad area. Only after the identification of four compatible layers the system will


**Figure 5.12:** *A girl paying a game session in the "Good or Bad?" interactive environment.*

perform the complete composition.[6]

### 5.2.1 How it Works

For the GoB game a database of couples of compositions belonging to the same musical genre is employed. The compositions are organized in four layers corresponding to similar instrumental parts, plus a fifth track where the four instrumental tracks and the melody are merged in a full audio rendering of the composition. Thus, each repertoire is composed of ten tracks, as shown in Figure 5.15. The couple of pieces play two different roles in the game. One is the reference piece, namely the piece to be recomposed, and the other is the antagonist, that is the disturbing piece. The game's logic assumes that listeners, after having heard the first track of the reference piece, are able to recognize if a second randomly selected track belongs to the reference or to the antagonist piece. Given the audio track random order, if the first player's choice $c$ is $1 \leq c \geq 4$, the first composition is the reference piece and the second is the antagonist, whereas if the choice is $5 \leq c \geq 8$, the second piece is the reference and the first the antagonist. The players' aim is to go through the track recognition process without mistakes until all the four tracks of the reference piece's are selected. The game's workflow is depicted in Figure 5.13. At the beginning of the game the players select a music repertoire. The eight tracks random order is generated and the reference and the antagonist piece is determined. At each next step of the first player, a new musical track is selected and played. If the track belongs to the reference piece it is accepted, otherwise it is discarded. But, before this happens, each time the second player must decide if

[6]An example of "Good or Bad?" utilization can be found at `https://youtu.be/XEliaFxKGeQ`

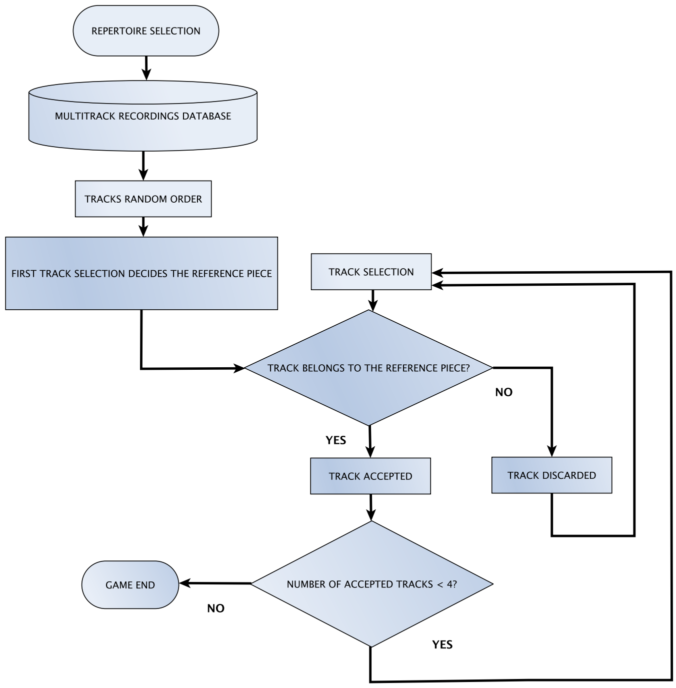



**Figure 5.13:** *The "Good or Bad?" game flowchart.*

the actual track is compatible or not with the ones previously selected, by stepping on the star or on the square of the right area. Then the system compares the track group membership with the second player's answer. If they are concordant a life is saved, otherwise is lost. Then the cycle begins again from the track selection onward, till the four compatible tracks of the reference piece are selected and evaluated. At this point the game is finished. If less or maximum three lives are consumed by the second player the game is won and the system performs the complete composition, otherwise the game is lost.

### 5.2.2 Related Work

Music listening is a widespread form of entertainment, greatly supported by the diffusion of cheap playback devices and mobile internet connection. In everyday listening the mere passive exposition to music supports many physiological and cognitive functions like implicit knowledge of complex harmonic and melodic structures (Rohrmeier and Rebuschat, 2012), sensorimotor answers to regular rhythmic inputs (Clayton et al., 2005), emotion arousal (Juslin and Laukka, 2004) and recovery enhancement in brain injured patients (Särkämö et al., 2008). In spite of this wide range of impact effects on the human brain and cognitive system, only a very little part of the interaction possibilities offered by musical listening are employed in the most popular contemporary entertainment applications like video games. Music is massively employed in video games, but mainly as a background element aimed at perceptual immersion, or characters con-notation (Vorderer and Bryant, 2012). Also in the most popular music video games, i.e. 'Guitar Hero',[7] user interaction is very often mediated by visual input, which plays a predominant role with respect to music content. Conversely, the GoB game logic is deeply grounded on musical listening. Ecological listening approaches derived from the principles of perceptual grouping,[8] provide elementary frameworks useful to outline some basic listening processes Bregman (1994). Notwithstanding, deeper level of musical understanding imply not only mere passive listening involvement, but also the arousal of more conscious perceptual mechanisms like attention and memory Sloboda (1985). This level of music listening and musical elements discrimination is the base upon which the GoB game is built. The GoB game can be placed in the framework of two main reference research fields: active music listening applications and video music games. Active music listening implies that listeners control somehow what they listen by performing meaningful actions on the musical content (Pachet, 1999). Section 5.2.2.1 provides an extensive review of some active listening applications and outlines the GoB game characteristics in this domain. To involve listeners in the high level of musical listening required by GoB, a formal game's system is employed as an engaging tool, including challenges and points achievement. This links the framework of the GoB game to the structure of music video games, which are analyzed in Section 5.2.2.2.

#### 5.2.2.1 Active music listening

Active music listening may be interpreted in many different ways. Volpe and Camurri (2011) distinguish between content and user centered active listening applications, whereas the first aim at providing a more creative approach to music content and the other depend strictly on user interaction or degree of users collaborativeness. To the first group belong MusicSpace (Pachet et al., 2000) where the user can manipulate the mixing of pre-recorded song audio tracks by changing the spatial position of the various instruments on a GUI interface, and Masataka Goto's various active listening applications aimed at song chorus detection (Goto, 2003), song navigation (Goto, 2007) and musical information display through a web interface (Goto et al., 2011). Interaction-based active listening relies on a single user gestures, as in the case of Orchestra Explorer (Camurri et al., 2007), or on many users coordinated interactions as in the Sync-in Team (Leman et al., 2009) game, where users have to synchronize their movements to a musical beat and achieve a rewarding score according to the degree of synchronization they reach. Also in Mappe per Affetti Erranti (Camurri et al., 2008), a project developed at Casa Paganini/Infomus (Genova, Italy), the musical output is linked to the interpreters collaborative behaviour, depending on the zone occupied by the dancers and on how their movements relate. Other experiments are reported in (Varni et al., 2012) where three applications based on users movements synchronization are described. As a matter of facts, in both application groups some action is performed on the musical content with the use of various techniques and materials. Notwithstanding, a dif-

[7] See `https://en.wikipedia.org/wiki/Guitar_Hero` for 'Guitar Hero' reference.

[8] The Gestalt theory provides a basic setup to understand how visual and audio perception works. It is based on six principles: proximity, similarity, closure, good continuation, common fate and good form, which explain how listeners organize a sequence of events in greater units which in a musical piece are organized at various levels of complexity.

ference may be drawn between applications which employ traditional computer graphical user interfaces and applications based on gesture or full body interactions. The possibility of involving gestures and users movements in human-computer interaction, has opened a wide set of possibilities for interface designers (Jacob et al., 2008), one of which is the richer interaction data availability which may help to design better and more intuitive interfaces for creative listening. Moreover, acting in the reality allows the possibility of considering the collaborative element which is regarded as an important design factor in many active listening applications. Table 5.3 and 5.4 report active listening applications subdivided in GUI and gestural/bodily/collaborative or social interaction applications.

| *Application/Main author* | *Music Material* | *Action* | *Task* |
|---|---|---|---|
| MusicSpace/Pachet | multitrack audio | mixing | post-processing |
| Smart MusicKiosk/ Goto | song audio track | chorus selection | archive navigation |
| Cindy/ Goto | song audio track | dance alignment | dance sequences navigation |
| Lyric Synchronizer/ Goto | song audio track/ metadata | lirycs alignment | lyrics navigation |
| INTER/Goto | song audio track | mixing/equalizing | post-processing |
| Musicream/ Goto | song audio track | searching for song similarities | archive navigation |
| Musicrainbow/Goto | song audio track | searching for artists similarities | archive navigation |
| Drumix/Goto | song audio track | adding tracks | composing |
| Songle/Goto | web song audio track platform | track analysis | music understanding |

**Table 5.3:** *Table of active music listening applications based on a graphical user interface.*

Both tables summarize the characteristics of active listening systems with respect to the music material employed, the action performed (i.e., the result of the processes applied on the audio file) and the application's task. Table 5.4 subdivides the interaction modality in gestural (single user), full body/social (bodily movements of a group of users) and gesture/social (coordinated gestures of two users). Though the original idea of active listening did not include any creative operation on the existing music material, "... we seek to create listening environments for existing music repertoires, rather than creating environments for composition or free musical exploration ..." (Pachet, 1999, p. 4), some active listening applications approach the idea of music creation. The Drumix application has the task of superposing a new generated drum track to the existing song tracks, whereas Mappe per Affetti Erranti, combining the emersion of instrumental parts with expressive interaction algorithms, can change very deeply the perception of the pre-recorded music material, approaching thus the manipulation levels of a real composition. The materials employed in active listening are mainly song audio tracks or multitrack audio. Song audio tracks undergo digital sound processes like acoustic features analysis and similarities detection (Goto, 2007) or digital filtering, whereas multitrack files are used to explore polyphonic compositions (Camurri et al., 2007), to mix tracks in a creative way as in MusicSpace, INTER, and

| *Application/Main author* | *Music Material* | *Action* | *Task* | *Interaction* |
|---|---|---|---|---|
| The Orchestra Explorer/Camurri | multitrack audio | mixing | exploration | gestural |
| Sound Scope Headphones/Hamanaka | multitrack audio | mixing | post-processing | gestural |
| Mappe per Affetti Erranti/Camurri | multitrack audio | selection/ expressive interaction | interpretation/ composition | full body/ collaborative |
| Sync-in-Team Game/Leman | regular pulse | beat alignment | performance synchronization | full body/ collaborative |
| Sync'n'Moog/Varni | song audio track | filtering | recomposition | gesture/ collaborative |
| Sync'n'Move/Varni | multitrack audio | adding tracks | recomposition | gesture/ collaborative |
| Sync'n'Mood/Varni | multitrack audio | synchronizing | recomposition | gesture/ collaborative |
| Good or Bad?/ Mandanici | multitrack audio | selection | recomposition | full body/ social |

**Table 5.4:** *Table of active music listening applications based on gesture or full body interaction.*

Sound Scope Headphones (Hamanaka and Lee, 2007). A further action allowed by multitrack recording is the song recomposition, which is made by adding the different musical layers corresponding to the various tracks one after the other till the composition's completion. This kind of music recomposition is the task of the Sync'n'Mov application (Varni et al., 2012), whereas Sync'n'Moog employs the spectral changes produced by a digital filter to allow the whole piece to emerge. Another technique has been experimented in the Sync'n'Mood application, where the recomposition happens through the synchronization of the piece's instrumental parts. In all these last three cases the user's reward is the compositions's full intelligibility obtained with users gesture alignment. A similar mechanism is employed in the Sync-in-Team game which is the only active listening application, except GoB itself, to be defined as a game.[9] As an active listening application GoB employs multitrack audio of songs or of other musical works to recompose the original piece. The tracks are selected by one player and then accepted or discarded by a second player. The game's aim is the music recomposition, obtained through body-operated, social interactions.[10]

[9]According to Juul's definition (Juul, 2011), a game is a formal system characterized by 6 main features: rules, variable and quantifiable outcome, valorization of outcomes, player effort, player attached to outcome and negotiable consequences.

[10]The word "social" instead of "collaborative" interaction is employed, meaning that here two players contribute to select and accept/discard the audio track in two different game phases, performing two different, separate actions, whereas "collaborative" expresses a simultaneous, aligned interaction aimed at obtaining a common result.

#### 5.2.2.2 Music video games

According to Pichlmair and Kayali (2007), music video games can be subdivided in three main categories: rhythm games (like Dance Dance Revolution, 1998 and Amplitude 2003), electronic instruments games (or first person shooters games like Rez or Elektroplancton, 2001) and music puzzles (the Ocarina of Time episode of The Legend of Zelda, 1998). In rhythm games the player is presented a series of visual rhythmic impulses based on preselected song collections[11]. The player's task is to repeat the sequence stepping on a dance pad or pressing buttons on a game controller. In these kind of games the song is considered as a mere rhythmic source which can be activated regardless of musical content, whereas the main cognitive impact on users is the high level of motor coordination necessary to play the game and not the evaluation of some music perceptual attribute. Moreover these games, thanks to the physical input modality, combine entertainment factors with health benefits, high level challenges and socialization issues (performances, competitions, dressing-up, interaction styles, etc. (Hoysniemi, 2006)). Contrary to rhythm games, the category of one-shooter games allows a greater freedom for the users, as the sound production depend on user interaction. Players have to follow some target appearing and wandering across the play space. When the target is hit, a sound is produced, building thus a strong, active relationship between game events and sound perception. The analysis of these relationships emphasizes important sonic interaction themes like acoustic ecology, auditory icons and sound localization (Grimshaw and Schott, 2007) that put games belonging to this group to a higher level of music cognition and participation. The third category of music puzzles builds upon the user's musical knowledge and requires listening skills, memory and musical elements knowledge and discrimination. A classical example of music puzzle appears in the Ocarina of Time episode of The legend of Zelda saga.[12] Here the melody pitches are mapped upon a controller input: the user is asked to play the right pitch sequence in order to match the required melodic pattern. The successful performance activates spells and environmental elements, allowing the game's progress. GoB belongs to the same category of music puzzle games. The game's second player, often supported by bystanders' opinion, has to solve a music puzzle to decide if the actual sound track fits the ones previously selected or not. This requires complex cognitive music processing, as described in Section 5.2.3.

### 5.2.3 "Good or Bad?" Music Information

The musical layers belonging to the same composition have strong horizontal and vertical relationships. Each layer displays musical elements which have the same timbre, metrical organization and musical phrasing, whereas each horizontal layer shares with the others the same musical tempo, accent occurrence relationships and key. These strong connections make listeners perceive the superimposed layers as a whole musical piece. An example of how these relationships work can be seen in Figure 5.14, where the tracks merging of two musical repertoires is com-

[11] See `http://www.ddrfreak.com/versions/listver.php` for an example of the Dance Dance Revolution song repertoire.

[12] See `https://en.wikipedia.org/wiki/The_Legend_of_Zelda:_Ocarina_of_Time` for The Legend of Zelda reference.

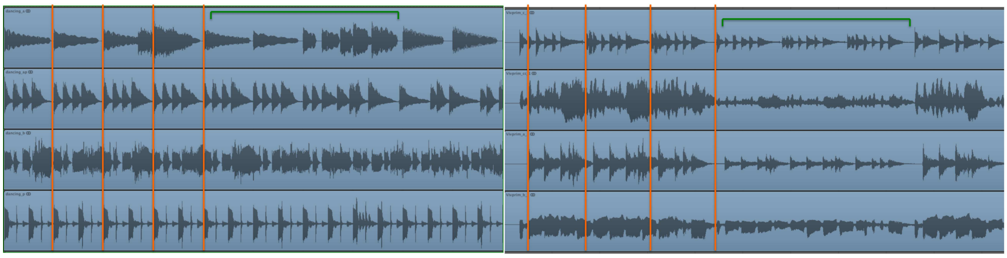


**Figure 5.14:** *Multitrack recording of Abbas's Dancing Queen and Vivaldi's La Primavera, Allegro. The red bars show the vertical metrical relationships, the green bar the musical phrasing.*

pared. The red vertical bars mark the effects of the shared musical tempo and meter, which produce the outlined occurrences. The green horizontal line marks the musical phrasing that organizes melodic and rhythmic patterns in longer unities. As the game's aim is to make listeners aware of the different parts which form a musical piece, sequential integration is employed to allocate sounds to the same layer and stream segregation to distinguish the characteristics of the different layers (Bregman, 1994). A visual analysis of the track recordings of the two repertoires shown in Figure 5.14 may outline some characteristics of the cognitive workload required to the game players: The first piece shows clearly differentiated patterns in the four tracks, whereas the second shows an higher degree of homogeneity. This suggests that stream segregation could be based mainly on rhythmic pattern detection in the first case and probably on timbrical or melodic contour contrasts in the second. Moreover, the second example shows an horizontal perceptual organization expressed by the lower profile of all the tracks in the zone marked by the green bar (an eco effect could produce such differences), whereas the first piece doesn't outline any feature of the kind. Thus, the above cited perceptual processes have to be regarded as general mechanisms which the listener must adapt to a great variety of situations.

### 5.2.4 The Geometric Interpretation of the Music Layers

The GoB game is based on multitrack recordings composed of four instrumental tracks (excluding the melody) plus a fifth complete track (instrumental tracks with melody). The reason of this choice is twofold: firstly, four is a good number to identify meaningful instrumental functions, typically percussion, bass and two accompaniment parts; secondly eight (four tracks for two compositions) is a good number to partition in an efficient way the applications' surface area reserved to the first player. To produce multitrack recordings fit for the GoB game, a database of musical compositions with available midi files downloaded from internet websites is employed.[13] The tracks are loaded on a musical editor and every repertoire is assigned the same group of virtual instruments to ensure a perceptual timbral similarity. Only the song chorus or the first musical phrase are selected for recording. The midi files come with a variable number of instrumental tracks, usually ranging from eight to twelve. Thus, a reduction is necessary to reach

[13]Popular free midi files sites are `http://www.midiworld.com/search/?q=pop` or `http://mididb.com/pop/`

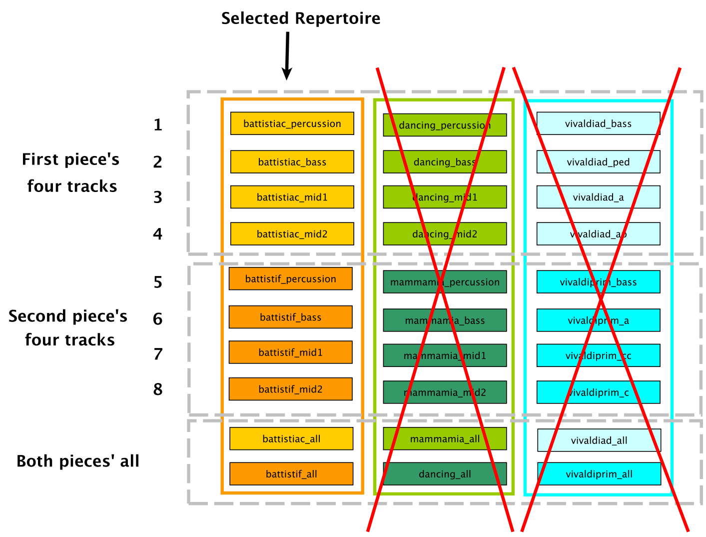



**Figure 5.15:** *The geometrical representation of the three repertoires employed in the "Good or Bad?" game subdivided in couples of pieces. The grey dashed lines mark the four instrumental tracks of each piece and the two complete piece's rendering tracks (all).*

the required number of four. The reduction process begins with the musical functions analysis, necessary to clarify the understanding of the various musical layers and to help the perception of the reciprocal differences. An example of musical functions analysis is reported in Figure 5.16. The "Dancing Queen" song comes with ten midi tracks: four tracks featuring the melody are skipped. Among the remaining six, the bass is recorded directly from the original track, whereas the two percussions are merged into one. A choice is required as far as concerns the remaining three accompaniment and piano tracks. The upper part contains steady chords, the part immediately below a chord rhythmical pattern with many rests, whereas the piano track features a constantly repeated chord pattern. Leaving the second track alone would mean including long silence zones in the recording, and this makes the track comparison required by the game rather difficult. Thus, track one and two are merged and track three is left alone with the aim to guarantee a perceptual balance with respect of sound continuity and rhythm clarity. As a result, a total of ten music tracks are obtained. The tracks represented in Figure 5.15 belong to three different repertoires. From the top there are the four instrumental tracks of the first piece, the four instrumental tracks of the second piece and the two tracks with the two pieces complete recording. This is the musical content representation upon which the whole game logic is based.

### 5.2.5 "Good or Bad?" Interface

The GoB playing area projected on the floor surface is divided in two zones marked by a black line, each reserved to one of the two players (see Figure 5.17). The left part is assigned to the first player, the right part to the second. In the left part the audio tracks are placed in four

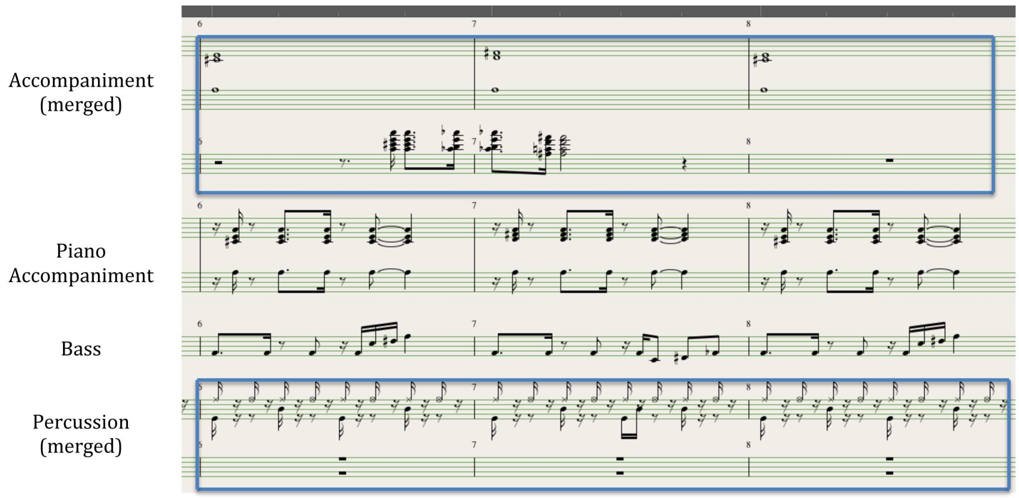



**Figure 5.16:** *Abba's "Dancing Queen" score excerpt with musical function reduction and merged tracks.*

stars (reference piece) and four squares (antagonist piece) put in a random order assigned by the system. In the right part only two buttons are visible: the good button (the star) and the bad button (the square).

### 5.2.6 "Good or Bad?" System Architecture

The Gob's system architecture is composed by three software modules, aimed at video analysis, game logic, and interactive graphics processing and audio files management. The video analysis and users' motion tracking is provided by the OpenPTrack software, which has already been described in Section 2.3.1.4. Game logic and interactive graphics have been programmed with Processing 3.0 beta,[14] whereas the audio files management is controlled by a Max/MSP patch. Processing and Max/MSP software modules communicate through the OSC protocol. The GoB interactive environment is equipped with the following devices:

a) a professional aluminium truss with dimension 4,00 x 3,00 meters where the motion sensors, the projector and the audio speakers are placed;

b) two Kinects v2 with adaptor for USB connection;

c) two Speakers Genelec 8030A;

d) one Projector Optoma X305ST placed on top of the above mentioned truss, which provides a projected floor of 4,00 x 3,00 meters with resolution 1024 * 768;

[14]Processing is a Java-based framework used to simplify development of visual contents, animations and games (`www.processing.org`)

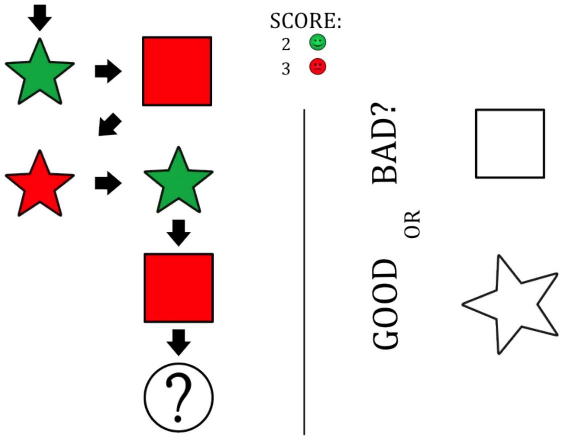



**Figure 5.17:** *The 'Good or Bad?' game projection of the playing area.*

e) three desktop computers: one that coordinates the tracking system and acquire data from a Kinect v2 (PC 1), one taht acquires tracking data from the other Kinect v2 device (PC 2) and one taht receives the information about tracked actors and executes the game, controlling the visual and audio feedback system (PC 3);

f) one Netgear GS208v2 network switch that connects all the PCs;

g) an optional white plastic sheet can be laid on the floor to increase the brightness and the definition of the projections.

The system is depicted in Figure 5.18. It works in every light condition and can track without occlusion many users at the same time.

#### 5.2.6.1 Audio files management

In the GoB game the listening process follows the game logic employed. Figure 5.19 shows a case of track selection process in a GoB match. To help listeners in the evaluation process, an audio stereo system with fixed audio volumes is employed. The left channel plays at a lower level and is used for the reference piece; the right channel plays at a louder level and is used for the track to be evaluated. At the beginning of the match the system produces a random sequence of the two pieces' music tracks. When the first track is selected (track 2 in the example), it is played in loop on the left audio channel, whereas the second track to be evaluated (track 5) is sent to the right channel. As it does not belong to the reference piece, the second track is muted. At the fourth step track 2 is playing on the left channel and track 1 is proposed on the right channel for evaluation. As track 1 belongs to the reference piece it is sent to the left channel and played in loop together with the first. Thus the reference tracks gather all together in growing number in the left channel at a lower volume, while the track proposed for evaluation stands alone in the right channel at a higher volume. Summarizing, in the example of Figure 5.19, the second,

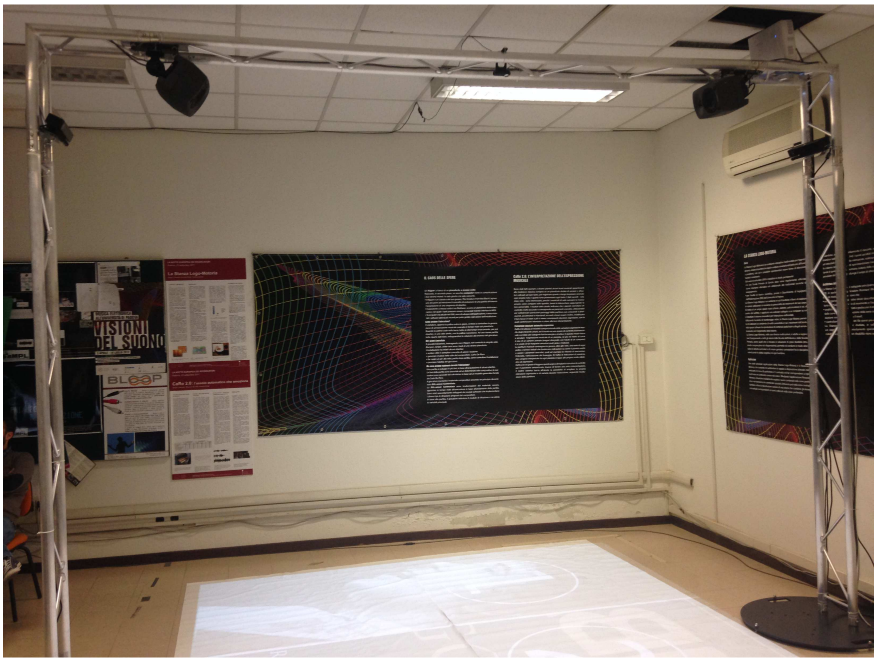



**Figure 5.18:** *The "Good or Bad?" interactive environment setup at the Sound and Music Computing Group Lab in Padova (Italy).*

third and fourth game step have the same listening condition, because the tracks are compared one by one. But, after the fourth step, a second reference track is added to the left channel, changing the perceptual balance of the evaluation. Actually, the tracks to be evaluated will be more easily assigned to their group, as the effect of the disturbing (or according) track will be more effective in this condition. This perceptual process is based on the well known figure-ground auditory segregation, that is the human hearing system's ability to detect an emerging acoustical figure against a background of other simultaneous elements (Teki et al., 2011). Pitch and time expectancies, belonging to listeners' previously acquired musical culture and memory, play a fundamental role in the track evaluation process, as shown by Dowling et al. (1987).

### 5.2.7 "Good or Bad?" Application Framework

The GoB application framework is depicted at Table 5.5. In the GoB game the users task is to recompose the reference piece through a two-player action that influences the listening conditions of a musical composition. The spatial positioning of interactive landmarks on the active surface is influenced by players interaction and different functions. In fact, the two players are given two separate portions of the interactive floor. upon which they behave in a completely different way. The $1^{st}$ player acts on the surface's portion where the musical information is represented. The random order of the spatial positioning of the musical layers well expresses the players' task, that is to recompose the reference piece through a series of listening tests whose result is to accept or to discard the random selected track. Thus, for the $1^{st}$ player the choice of where

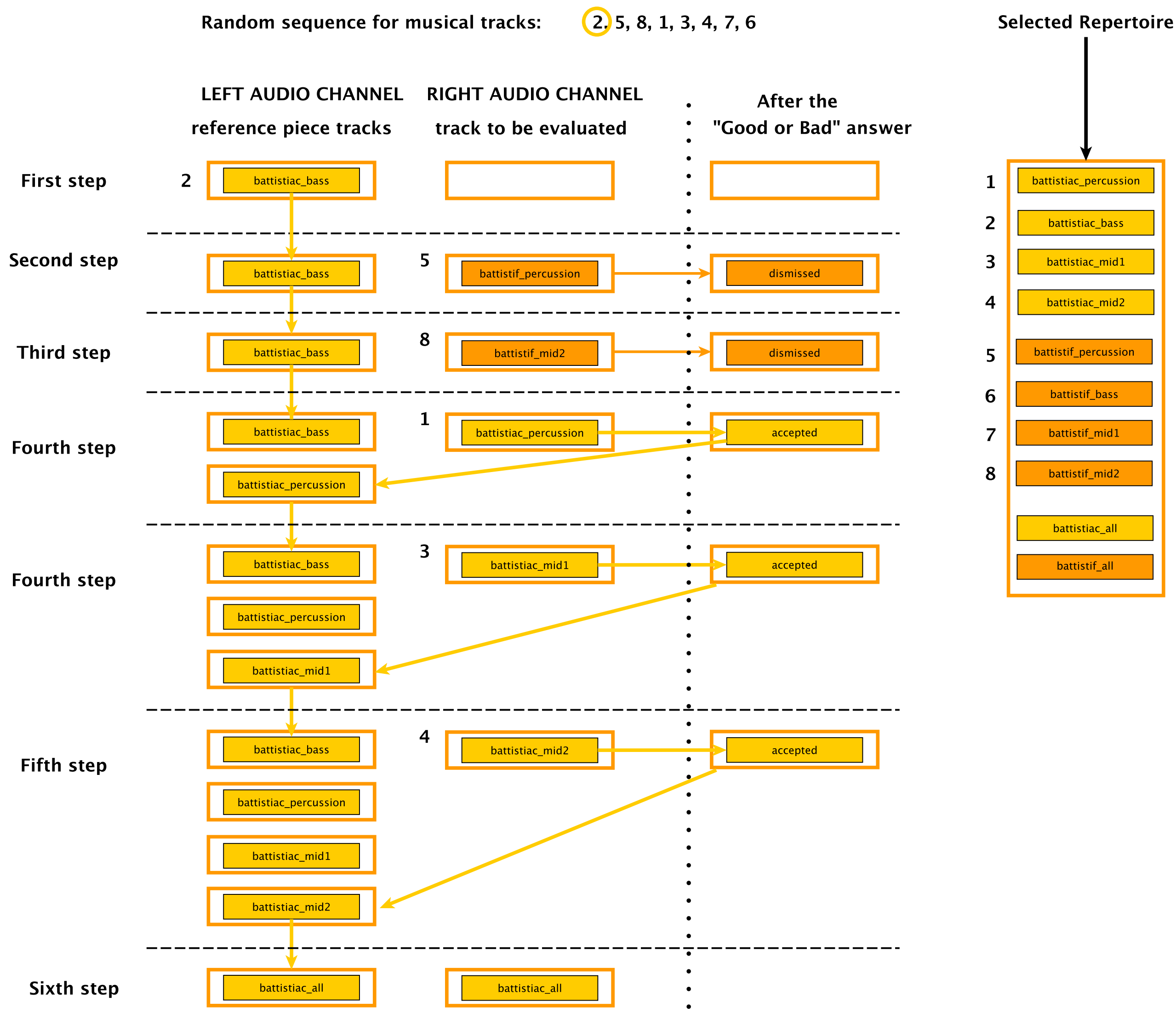



**Figure 5.19:** *A case of track selection process in a "Good or Bad?" match.*

<table>
<tr><th colspan="4">”GOOD OR BAD?” APPLICATION FRAMEWORK</th></tr>
<tr><td colspan="2">Aim</td><td colspan="2">Music Recomposition</td></tr>
<tr><td rowspan="4">Interface</td><td>Musical Information</td><td colspan="2">Instrumental layers</td></tr>
<tr><td>Geometric Interpretation</td><td colspan="2">A stack of four superimposed instrumental layers + a complete piece rendering for each composition</td></tr>
<tr><td rowspan="2">Spatial Positioning</td><td>$1^{st}$player</td><td>$2^{nd}$player</td></tr>
<tr><td>Four stars and four squares buttons in two rows in random order</td><td>One star and one square botton (good or bad)</td></tr>
<tr><td rowspan="4">Interaction</td><td rowspan="2">Where to move</td><td rowspan="2">Contiguous positions</td><td>Before the choice: towards the good or the bad button</td></tr>
<tr><td>After the choice: outside the active area</td></tr>
<tr><td rowspan="2">When to move</td><td>$1^{st}$ time:<br>wait until the available position appear</td><td>When ready for decision</td></tr>
<tr><td>$2^{nd}$ time and on:<br>after the $2^{nd}$ player has made her/his choice</td><td>Immediately after the choice</td></tr>
</table>

**Table 5.5:** *The table shows the “Good or Bad?” aim, musical information, geometrical interpretation and spatial positioning of interactive landmarks subdivided for each player. The interaction area explains how players behave with respect to their different functions.*

to move is completely free, nor in any way it may affect the fate of the match. On the contrary, the timing of her/his movements is constrained mainly by the $2^{nd}$ player’s times of decision and, only after the first selection, by the system itself. At the beginning of the match the $1^{st}$ player actions are blocked in a ten seconds listening time to allow both players to well understand the musical qualities of the first track that is the first hint of the reference piece. Thus, the $1^{st}$ player has to wait until the ten seconds expire and the graphics with the available positions appear on the floor. From the second time on the available positions will appear as soon as the $2^{nd}$player has made her/his choice. The $2^{nd}$ player has only three positions available: the star button (good), the square button (bad) and the position outside the active area that must be reached immediately after the decision has been made. After the $1^{st}$ player has selected the audio file, the $2^{nd}$ player has to choose if it belongs to the reference piece or not. The time s/he takes for this decision depends both on $2^{nd}$ player’s ability and on the listening condition, as discussed in Section 5.2.6.1. Moreover, as the game’s interface allows the participation of all the people around, also the time for a collective discussion must be taken in account. Summarizing, players are closely dependent on one another in their movements. The $1^{st}$player plays the role of a localized selector (a step on an interactive landmark) whereas the $2^{nd}$player confirms or discards the choice. This is an example of coordinated selection as described in Section 3.2.1. These actions are operated upon the musical material but the interaction does not depend on music qualities, like in Harmonic Walk. Thus, GoB is not a play-acting game but is a challenge game, because users are required to solve a problem employing their musical listening abilities.

## 5.3 Case Study 3: Jazz Improvisation, an Active Listening Application.

Jazz Improvisation is an active listening application based on composition techniques employed in jazz and in many other popular music genres. Musicians start from a shared musical mode - which provides a pool of available notes - and from a common musical meter. They improvise their instrumental musical patterns and combine this material in various ways. As the patterns can be freely superimposed, the application's surface is subdivided into nine zones without any particular element relationship. When a zone is occupied for the first time, the music pattern is played, but, if the user returns on the same zone, the corresponding audio file is muted. This creates a dynamic relationship between the user's path and the composition state, which changes every time the user moves.[15]

### 5.3.1 Jazz Improvisation Musical Information

Modal music is very common in ethnic music, ancient music and modern jazz. Though the term usually refers to the eight ancient gregorian modes, the widespread utilisation of modal music in $19^{th}$ and $20^{th}$ century has extended the range of used modes to a series of different musical scales. As an example, Béla Barók has employed many ethnic music modes in his compositions.[16] The lack of the strong contrasting relationships that characterise tonal music (e.g., the dissonance between the $4^{th}$ and $7^{th}$ degree in the major and minor scales) makes modal music much more ductile with respect to tonal music. Particularly, modal music allows slow-moving harmonic rhythm sections, where long-lasting single chords may be the basis for instrumental improvisation. Pedal points, drones,[17] and "ostinatos"[18] are also characteristic features of modal music. This kind of musical structures are suitable of successive stratifications of various instrumental parts, as it happens in jazz improvisations, in didactic music (Keetman and Orff, 1950 - 1954) and in popular music.

### 5.3.2 Jazz Improvisation Interface

For the Jazz Improvisation, nine musical parts have been composed using a common mode, musical meter and ostinato patterns. As the nine musical layers are independent one from the other and may be recomposed in a free way, they have been scattered on nine interactive landmarks obtained through an equal distribution of the active application's active surface. It has been employed as an interactive wall projection to help user to realize their actual position and active layers status while moving on the active floor, whereas the interactive landmark position has

[15] A Jazz Improvisation virtual performance `https://youtu.be/uI3trfpPakU`

[16] Béla Bartók (1881 - 1945) is an Hungarian composer who devoted a considerable part of his musical activity in the research and analytical study of folk music.

[17] Pedal points are long, sustained notes usually in the bass, upon which also dissonant melodic parts or harmonies can be superimposed. Drones (or "bordone") have similar functions but, instead of only one note as in the pedal point, it may be formed by a fifth or by a whole chord.

[18] An ostinato is a musical phrase that is continuously repeated in the same part.

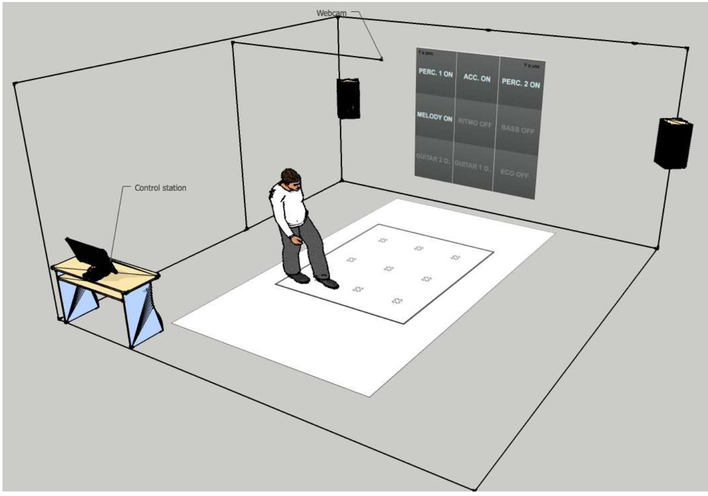



**Figure 5.20:** *The Jazz Improvisation environment with a wall projection showing the playing zones activated after the user's step.*

been marked with white crosses on the floor, as depicted in Figure 5.20 where the wall projection shows the playing zones already touched by the user. The Jazz Improvisation system architecture is similar to the Harmonic Walk system already described in Section 5.1.7.

### 5.3.3 The Jazz Improvisation Application Framework

The Jazz Improvisation framework depicted at Table 5.6 shows the same musical information of the GoB application. But the geometric interpretation shows the difference between the two: in the GoB the layers originate from a complete composition which has been sliced in four layers plus the melodic part. Thus the geometric representation is a stack where all the layers have a precise place and whose simultaneous performance generates the complete composition. In the Jazz Improvisation the nine layers do not belong to a complete composition but may be freely superimposed without any claim of completeness. Thus, their geometric representation is a series of interactive landmarks freely arranged on the application's surface. The user can begin her/his exploration of the interactive landmarks from any accessible point along the borders. However, the possibilities to reach the landmarks depend on the path the user choses, as not all the landmarks can be reached from any point. As depicted in Figure 5.21, the path to activate a precise set of landmarks must be carefully planned by the user who has to find her/his way as if moving in a musical labyrinth. In this case the problem of where to move may have two solutions: to make a random exploration of the environment whatever the musical result, or plan a precise path through the environment in search of a precise musical result. The choice of when to move depends only on the listening time the user needs to understand the music layer s/he is activating and on how s/he wants to use this information. For instance, the user could activate a percussion landmark only for a limited amount of time and than mute it by stepping again. Thus, the timing of the movements depend on the user listening abilities and creativity in designing the

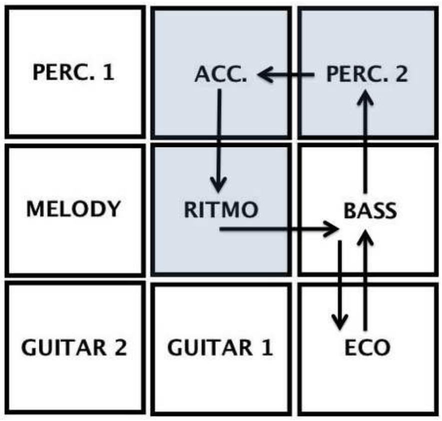



**Figure 5.21:** *A possible path on the Jazz Improvisation surface. The blue interactive landmarks are the playing ones because they have been touched only once by the user, whereas "eco" and "bass" are muted after the second user's step.*

| **JAZZ IMPROVISATION APPLICATION FRAMEWORK** | | |
|---|---|---|
| **Aim** | | Music Recomposition |
| **Interface** | *Musical Information* | Instrumental layers |
| | *Geometric Interpretation* | Nine free superimposed instrumental layers |
| | *Spatial Positioning* | Subdivision of the rectangular active floor area in nine rectangular interactive landmarks |
| **Interaction** | *Where to move* | $1^{st}$ time: begin from any zone |
| | | $2^{nd}$ time on: move to one of the two contiguous positions |
| | *When to move* | When the need of adding/muting a music layer is felt |

**Table 5.6:** *The table shows Jazz Improvisation aim, musical information, geometrical interpretation and spatial positioning of interactive landmarks.*

musical environment s/he prefers. As the application's aim is a free music recomposition based on environment exploration and musical discovery and listening, it can be assigned to the inquiry games category discussed in Section 3.2.3.

# Chapter 6

# Three-dimensional Spaces Music Applications

In this Chapter two music applications based on three-dimensional spaces are presented. Both applications deal with interactive music composition and performance expressive interaction. Disembodied Voices employs a user-centered three-dimensional space where the user acts in a hemispherical space in the range of her/his stretched arms. Following the conductor's model, the user performs impulsive input gestures to progress an interactive algorithmic composition and continuous input gestures to render an expressive performance. "Hand Composer" has similar functions as Disembodied Voices, but employs a smaller sensor-centered space. Both application belong to the category of "no touch" instruments, where the user employs kinesthesia and proprioception to direct her/his gestures towards open-space, invisible active regions and thresholds.

## 6.1 Case Study 4: Disembodied Voices, a Conducting System

Disembodied Voices is an interactive environment designed for an expressive, gesture-based musical performance. The motion sensor Kinect, placed in front of the performer, provides the computer with the three-dimensional space coordinates of the two hands. The application is designed according to the metaphor of the choir director: the performer, through gestures, is able to run a score and to produce a real-time expressive interpretation. The software interprets the gestural data and controls articulated events to be sung and expressively performed by a virtual choir. Hence the name of the application: you follow the conductor's gestures, hear the voices but do not see any singer. The system also provides a display of motion data, a visualization of the part of the score performed at that time, and a representation of the musical result processed by the compositional algorithms.[1]

[1]A Kinect Conductor video may be found at `https://youtu.be/oyf7GrMMrL8`.

### 6.1.1 How it Works

Disembodied voices is a system where a conductor plays a score written for music composition algorithms. According to Mulder (2000, p.316) in the musical tradition there are two forms of musical rendering: the conducting and the performance. The first is a symbolic gesticulation that aims at controlling the musical structures; the second consists in the manipulation of the control surfaces of the musical instruments aiming at an expressive sound production. In Disembodied Voices the conductor performs both functions. The environment is focused on the user's body positioned at the center of a virtual hemisphere, delimited by a radius corresponding to the length of her/his stretched arms. The user's right hand movements data produce discrete inputs, or "buttons" (Verplank, 2003), that are employed for the progress of an algorithmic composition score written according to the model of the so-called Ligeti's "micropolyphony" (Bernard, 1994). The left hand movements produce a continuous data flow, or "handles", employed for expressive interaction. Thus, the conductor acts on the musical structure creating interactively the timing succession of the different sections of the composition. In the same time, using another gestural repertoire, s/he interprets the score moving on the imaginary control surfaces relative to the interpretative parameters.

### 6.1.2 Related Work

Interactive conducting dates back to Mathews's Radio Baton (Mathews, 1991), where control signals were produced employing magnetic capacitance to effect sound production, in a way similar to the Theremin.[2] In more recent times the Digital Baton (Marrin and Paradiso, 1997) developed a similar idea employing a sensorized handheld device capable of detecting beats and movements acceleration for sound expressive interaction. Many other projects like Personal Orchestra, (Borchers et al., 2004) UBS Virtual Maestro (Nakra et al., 2009) and others, (Toh et al., 2013) and (Maes et al., 2013), were proposed with the main goal to control the musical tempo and sound dynamics. However, all these systems are inspired by a rather traditional idea of the musical conductor, who interacts with the musical result only through the beat and dynamics control. Digital sound processing techniques, as well as algorithmic composition, offer much wider possibilities of controlling music production in real time than the simple beat speed or sound volume. Moreover, Disembodied Voices employs the three-dimensional space all around the conductor as an active space, where imaginary thresholds delimit active regions for digital sound processing. Thus, Disembodied Voices can be considered as a no-touch instrument, because the user acts with her/his hands in the air, crossing imaginary region borders and moving inside them for expressive sound interaction. These are the reasons why the most important Disembodied Voices's reference is the Theremin. The Theremin uses a simple capacitance measurement to sense the proximity of the player's hand, thus providing the creation of a sensible field that is the instrument's no touch interface. Also if the sensing techniques employed by the two systems are very different, the playing technique of the Theremin is incredibly close to that employed in Disembodied Voices. The Theremin control space has two dimensions and is completely imaginary. Playing the instrument requires a highly developed kinesthetic sense: it

[2] The Theremin is the first electronic instrument, invented in Russia in 1920 by Léon Theremin.

is not only necessary to recognize the position in space (proprioception), but also to have the ability to understand the extent, direction and weight of the movements (kinesthesia). According to Billinghurst and Buxton (2011) the success of the Theremin would be right in this direct relationship between hands position in the control space and the continuous sound feedback that allows the player to build his own mental map for playing the instrument. Another no-touch instrument is the Sensor Chair[3] that is interesting precisely because it introduces for the first time the concept of a three-dimensional control surface. The Sensor Chair is a device that measures the hands and feet position and motion of a seated occupant. It was developed for MIT Digital Expression Conference in October of 1994 and it has been used as a one of the performance instruments in Tod Machover's Brain Opera.[4] A copper plate affixed to the top of the chair cushion is a transmitting antenna being driven at roughly 70 kHz. When a person is seated in the chair, s/he effectively becomes an extension of this antenna; her/his body acts as a conductor which is capacitively coupled with the transmitter plate. Four sensors provide the $xy$ plane position for the hand as well as $z$ position of the hand's distance from the sensor plane. This coordinates data have been used to launch a sound and adjust its volume ($xy$) and to change its timbral characteristics ($z$), or to divide the $xy$ plane into many zones which contain different sounds (Paradiso and Gershenfeld, 1996).

### 6.1.3 Disembodied Voices Musical Information

In Disembodied Voices the conductor employs hand movements to communicate with the musical system. Movement analysis (Marrin and Picard, 1998) as well as academic teaching (Rudolf and Stern, 1994) subdivide the role of the two hands. In general the right executes musical cues while the left is devoted to iconics, metaphorics and dynamics. Both hands and eyes perform deictic gestures, which express spatial and temporal features of the musical events. The right hand and eyes give the attacks while the left hand alludes to the expressive content. The entire conductor's body communicates to the musicians the phrasing, whereas specific gestures can mimic instructions on how to perform particular accented notes or appoggiaturas or sforzatos (Rudolf and Stern, 1994). According to Humphries (1968), in conducting it is also very important to know what needs to be communicated, in what order and how. Conducting implies that the director is aware of the formal, harmonic, rhythmic and expressive structure of the music. A music composition may be considered as a series of ”expressive units” that are emotional, time-limited entities, which include the following features: duration, dynamics, rhythmic organization, articulation, continuity (or less) to the next unit, vitality (crescendo, diminuendo, accelerando) and so on. The conductor's task is to identify them, to outline their differences and to convey attention on them, building thus a “.... timeline of rhythmic attention shifts” (Humphries, 1968, p.2). These important gestural features as well as the hands role subdivision have been considered to build the computational model of the musical conductor. As Disembodied Voices interprets the result of music composition algorithms that are not inspired by traditional rhythmic musical models, no information about beat detection or musical meter pattern changing has been considered.

[3]Sensor Chair `http://web.media.mit.edu/~joep/TTT.BO/chair.html`
[4]Brain Opera `http://park.org/Events/BrainOpera/`

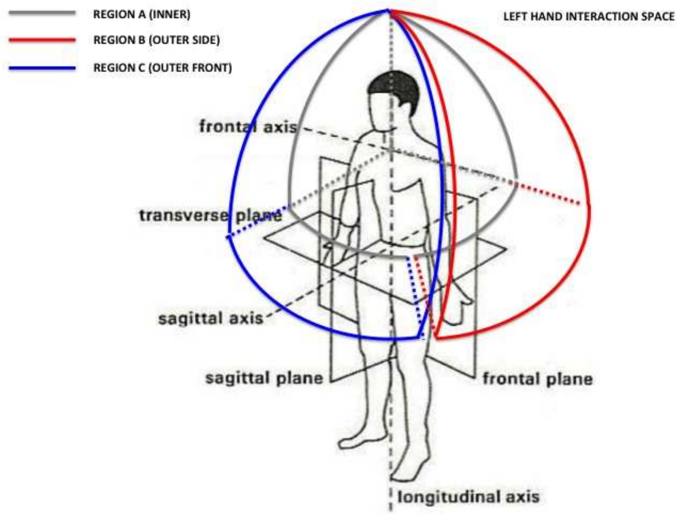



**Figure 6.1:** *The left hand interactive three-dimensional space of the Disembodied Voices. The space is subdivided in three zones: the grey inner region, the blue outer front region and the red outer side region.*

### 6.1.4 Disembodied Voices Geometrical Interpretation

The geometrical interpretation of the conductor's interaction space is a hemisphere with the center at the basis of the neck of the conductor and the diameter corresponding approximately to the two stretched arms length. The base of the hemisphere lies in the transverse plane, while the vertical radius is along the sagittal plane. Following the conductors interaction model, the hemisphere is subdivided in two parts, one for the right and the other for the left hand. The left hand interaction region is depicted in Figure 6.1, where three different spherical portions are shaped. This is a subjective interpretation of the conductor's interaction space aimed at enriching the expressive possibilities of her/his gestures. The actual conductor employs hand gestures, glances, eyebrow movements and many mimics to communicate her/his intentions to performers. All these expressive possibilities cannot be transferred in the actual computational model. Thus, to counterbalance the loss, this system of hemispherical regions is proposed. As soon as the hand enters the outer hemispherical region two digital sound processes are triggered, one in the outer front region and the other in the outer side region.

### 6.1.5 Disembodied Voices Interface

The three-dimensional space of the hemisphere is mapped through the spherical-polar coordinate system, where the arm's length is represented by $r$ (the radius), the lateral movements angle by $\phi$ (the azimuth) and the elevation angle by $\theta$ (the zenith). Therefore, the inner hemispherical region is determined by a threshold along the radius length, the outer regions are delimited by azimuth

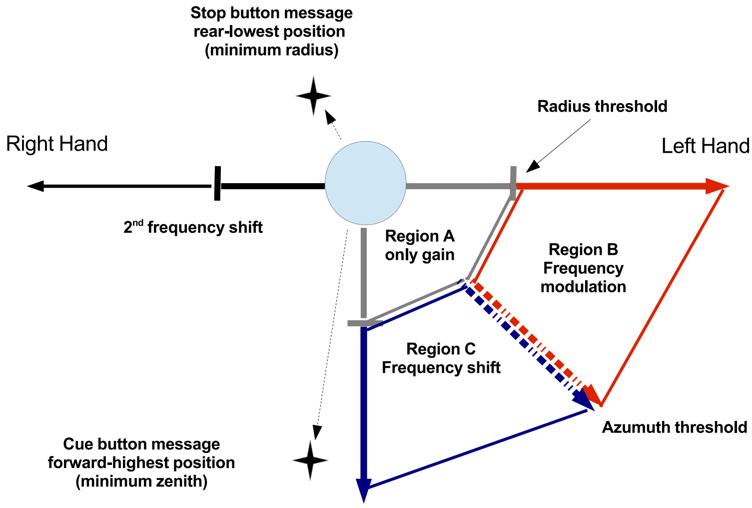



**Figure 6.2:** *Disembodied Voices control system.*

and radius thresholds, whereas the dynamic level depends on the elevation angle. The whole Disembodied Voices control system is depicted in Figure 6.2, where the active regions laying on the sagittal plane are seen from above the user's head. When the left hand moves upwards or downwards causes a variation of the value of the zenith, which effects the dynamics of the sound according to the common metaphor low-piano and high-forte (region A). An imaginary threshold in front of the performer is drawn depending on the radius value. As soon as the arm exceeds the threshold another subspace dedicated to the control of a spectrum frequency shift effect is activated (region C). The frequency shift value is minimal near the threshold and grows gradually pushing the hand forward. By adding another threshold on the azimuth, a subspace dedicated to the control of a modulation effect is created (region B). The right hand is mainly devoted to button input production. Two thresholds near the zenith minimum and radius minimum are set to obtain respectively a cue button and a stop button. A threshold on the right hand radius activates a second frequency shift effect.

### 6.1.6 Disembodied Voices System Architecture

The software structure of Disembodied Voices is divided into three functional blocks. The first block is the application Synapse which processes visual data from the Kinect sensor.[5] The second and third blocks are two distinct Max/MSP patches, one for data calibration and sound processing, the other for music composition algorithms and process visualization. The communication between Synapse and Max as well as between the two Max patches is done via the OSC proto-

[5]Synapse `http://synapsekinect.tumblr.com/`

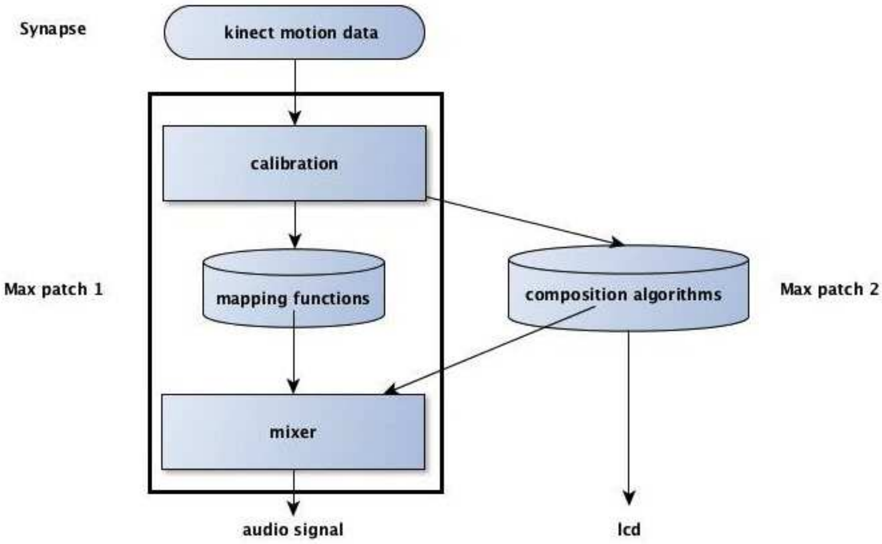



**Figure 6.3:** *Disembodied Voices's system architecture. The Kinect data are processed by the Synapse application and then sent to the Max patch 1 for data calibration and data mapping. Button inputs are sent to the Max patch 2 for algorithmic composition production and process visualization.*

col. Synapse collects the skeleton bone joints data of the performer and sends them to the other applications through the OSC protocol. Also if Synapse can process the tracking of many joints (hand, elbow, foot, knee head and torso), in the present application's implementation only the two hands data are employed. Figure 6.3 shows the overall flowchart of Disembodied Voices.

#### 6.1.6.1 Calibration of motion data

The motion data coming from Synapse are sorted into two data streams corresponding to the right and to the left hand. These data require further calibration to measure the maximum and minimum extension of the arms of the various users. This is automatically processed on the fly, allowing the user to speed up the calibration operation. After taking the psi pose[6] it is enough to hold arms forward, upward and sideways. The system records the maximum and minimum extension of the arms and use them to scale the movement data.

#### 6.1.6.2 Data mapping

Disembodied Voices manages two types of data: continuous controls for expressive interaction and discrete controls for triggering events (cues). The latter are performed by the right hand when reaching the extreme forward-highest position (minimum zenith) for cues attacks, and

[6]The so-called psi pose is a particular position, similar to the greek letter $\psi$ which allows Kinect to interpret the user's dimensions.

| action | movement | radius | az.th | zenith |
|---|---|---|---|---|
| **gain (a)** | L.H. cont. | - | - | y |
| **F.s. 1 (c)** | L.H. cont. | >25% | <25% | - |
| **mod. (b)** | L.H. cont. | >25% | >25% | - |
| **cues** | R.H. front imp. | max | 0 | <10% |
| **stop** | R.H. rear imp. | <20% | 0 | max |
| **F.s. 2** | R.H. cont. | - | x | - |

**Table 6.1:** *Summary table of hands movements mappings in Disembodied Voices.*

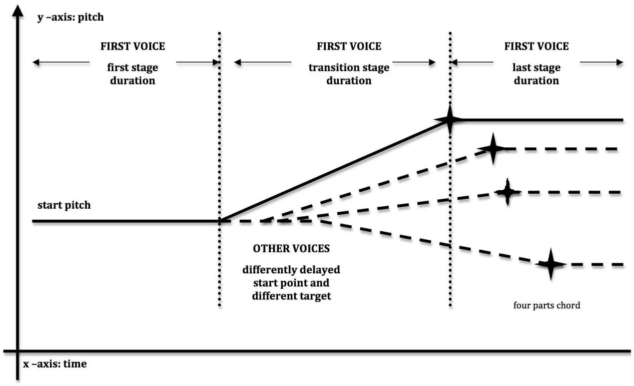



**Figure 6.4:** *Visualization of the breakpoint function of four delayed voices in the Ligeti's micropolyphony algorithm.*

the extreme rear-lower position (minimum radius) for stopping the performance. Continuous controls are mainly associated with the movements of the left hand. They are divided into: continuous signals, active all over the field range of mapped data (zenith for the dynamics) and continuous signals active only after passing a threshold (azimuth for modulation and radius for frequency shift). A second frequency shifter has been added in order to create beatings between the transposed sounds. This effect is activated by the right hand while moving into a dedicated subspace defined by another radius threshold. All hand movements mappings are summarized in Table 6.1. Since the gain is a value that interacts directly with the audio, its output is continuous and corresponds to a value ranging from -40 dB to 0 dB. Conversely, the other audio parameters are active only when the user's arms exceed the threshold corresponding to the 25% of the overall range value. The limits are 0-80 Hz for the modulation frequency, and 0-600 Hz for the frequency shift. When the right hand goes below the 10% of the zenith value an impulse is sent to a counter which outputs a specific cue number. Stop is sent when the right hand position goes below the 20% of the radius value. All button and handle messages are sent to the second patch via OSC.

#### 6.1.6.3 The music composition algorithms

Cue and stop messages are destined to the music composition algorithms. These are inspired to the model of Ligeti's micropolyphony, which is usually generated by a canonic compositional technique (Jarvlepp, 1982). However, in this algorithmic model micropolyphony has been considered as a chaotic mass movement of a given number of voices from one or more start pitches towards one or many target pitches. The movement is divided into three stages and can be represented by a breakpoint function which requires at least three couples of data: start pitch/first stage duration, target pitch/transition stage duration, target pitch/last stage duration, as can be seen in Figure 6.4. The start pitch and the target pitch may be the same for all the voices. If they differ, it means that the path of each voice is different, generating thus an interesting polyphonic texture. In addition, the starting point of the function can be slightly delayed by a duration value different from voice to voice. This difference produces a kind of automatic canon whose density can be controlled using appropriate delay values. The polyphony management in Max/MSP is realized by using the *poly∼* object, which easily duplicates voices instances and facilitates their control. Each *poly∼* object performs only one note. The objects are grouped in four sections named - according to the traditional choir voice division - sopranos, altos, tenors and basses, and are marked by an index. Each cue message contains a label which addresses one of the four vocal sections. The sequence of messages is fixed in advance and depends on the features contained in the musical score. After receiving the message, the composition algorithm produces a package of MIDI data that is sent to a sampler with synthesized voices.

#### 6.1.6.4 Digital sound processing

The audio signal coming out of the sampler is processed by a ring modulator and a frequency shifter. The modulation frequency, when active, grows progressively from 0 to 80 Hz (high threshold). It feeds a complex ring modulation that allows to translate the spectrum of the carrier signal around a modulating frequency as a normal ring modulation but with the possibility to select whether or not to include the lower spectral band. The frequency shift is obtained by real time analysis and resynthesis of the sound through FFT (Fast Fourier Transform) and IFFT (Inverse Fast Fourier Transform). The forward movement of the left hand shifts the spectrum by a value that ranges from a minimum of 0 to a maximum of 600 Hz. The effect distorts the original voice spectrum and the outcome sounds more aggressive. There is also the possibility of triggering a second frequency shift affected with continuous movements of the right hand beyond the appointed radius threshold. The difference between the two frequencies causes harmonics overlapping and, consequently, beats production.

### 6.1.7 Disembodied Voices Application Framework

Also if the conductor in Disembodied Voices acts both as a composer and as a performer, the application's musical information and geometrical interpretation concerns only the expressive interaction function. The Disembodied Voices composition algorithms structure is not open to interaction. Thus, no geometric representation is necessary, because the performer can only

| DISEMBODIED VOICES APPLICATION FRAMEWORK | | | |
|---|---|---|---|
| **Aim** | | Expressive performance of algorithmic compositon | |
| **Interface** | *Musical Information* | The musical conductor model | |
| | *Geometric Interpretation* | User-centered emispherical space | |
| | *Spatial Positioning* | **Interactive score production** | **Expressive performance** |
| | | Cue button at the forward-highest<br>Stop button at the rear-lowest | Emispherical regions partitioning |
| **Interaction** | *Where to move* | To the cue/stop button position | *Score alignment phase:*<br>to the appointed space region |
| | | | *Expressive alignment phase:*<br>to the best sounding position |
| | *When to move* | When the change to the<br>next section is felt necessary | *Score alignment phase:*<br>in due time with musical score directions |
| | | | *Expressive alignment phase:*<br>whenever an expressive change is needed |

**Table 6.2:** *The table shows the Disembodied Voices aim, musical information and geometrical interpretation. From the spatial positioning on, the table is subdivided following the two application's tasks: interactive score production and expressive performance.*

decide the duration of the various sections of the score but cannot change the musical structure of the event. Thus, the difference between interactive score production and expressive performance features is present only at the spatial positioning level, where one button for cue attack and stop are positioned at the forward-highest and rear-lowest positions respectively. The interactive regions of the Disembodied Voices interface are delimited by some thresholds positioned on the sagittal (azimuth), frontal (radius) and longitudinal plane (zenith). Unlike the active floor applications interfaces, where the user's attention focuses on the area of the various interactive landmarks, in the three-dimensional space the attention focuses more on the thresholds than on the regions. The main reason is that in Disembodied Voices's three-dimensional space the active regions are not visible to the user. The characteristics of musical performance employing these invisible regions are explained in the next Section 6.1.7.1.

#### 6.1.7.1 Performing on a no-touch Instrument

An important feedback when playing a musical instrument is the haptic sensation related to the mechanics of the instrument. This feeling is divided into: tactile (pressure, surface curvature, orientation, moisture, friction, etc.) and proprioceptive and kinesthetic sensations (Rovan and Hayward, 2000). They are so important to be considered as part of the instrumental performance. For example, they allow the players to orient themselves in taking positions, to improve tuning based on simple auditory feedback, to facilitate articulation, phrasing and thus expressiveness, to improve music pieces memorization (via the retention of certain situations resulting from haptic gestural sequences), to perform complex compositions or to read them at sight focusing the eye

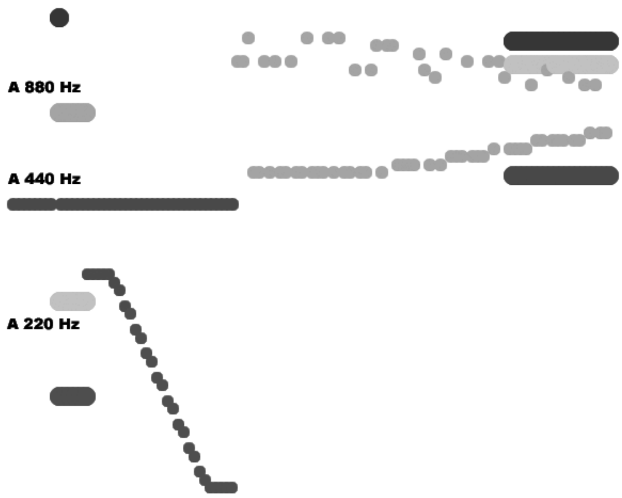



**Figure 6.5:** *Visualization of an example of the musical result of Disembodied Voices compositional algorithms. In particular, starting from the left, an A (440 Hz) pedal, four short A at different octaves, and a slide towards a very low A.*

on the score rather than on the instrument, etc. The characteristic of the no-touch instrument is to communicate with the computer through gestures without being in direct contact with the control surface. As a consequence, the haptic feedback in these performances is totally lacking. Rovan and Hayward (2000, p.300-301) accurately describe the performance process under these conditions, trough the following steps:

1. the performer, who is more or less forced by the system, must have in his/her mind an executive idea (intention)

2. then s/he must run the gesture to perform the intention. Through the control surface (which is imaginary and of different sizes and shapes) the gestural data become sound

3. evaluate her/his action through the vision and the kinesthetic sense

4. hear the resulting sound

5. amend her/his acts through the vision and proprioception

6. change her/his intention by appreciating the resulting sounds

7. and so on .......

It's obvious that in this scenario the lack of haptic feedback increases the importance of the visual and proprioceptive channels. The first Max/MSP patch is not only used for gestural data processing but also to produce a visual feedback of the interaction. For this purpose various GUI objects of different sizes and colors have been used to display the current state of the interaction in order to make it clear even far away from the screen. The second patch produces a visual


**Figure 6.6:** *Nine stages of a Disembodied Voices score noted in summary form for performer's reminder. Pitch events (la lungo, long A), dynamics (p, mp, etc.), ring modulation effects (MP, P, etc.), frequency shift ($\wedge$) and second frequency shift ($+$) are displayed.*

feedback of the result of the compositional algorithms by using a *lcd* object where small spheres display the sounds produced by these algorithms. On the $x$ axis is represented the time, on the $y$ axis the pitch (see Figure 6.5). The problem of all these visual feedbacks is that they are produced retrospectively, just after the action has been concluded or the position has been reached. They have only a confirmatory function of the intention once it has been properly completed. Unfortunately, they do not have any predictive function with respect to a possible mistake. In the case of Disembodied voices a real time three-dimensional display of the interpreter's avatar could be useful, as well as the graphics of the thresholds in the three-dimensional control space. The Max/MSP patch 1 also displays instructions for the performance of each cue. These notes act as a reminder for the conductor and include current event description as well as hints about the expressive interaction to be performed, as shown at Figure 6.6.

#### 6.1.7.2 The Disembodied Voices interaction paradigm

The interaction paradigm of Disembodied Voices is subdivided in two sections depending on the different performances tasks (interactive score production and expressive performance). Information about both tasks is contained in a score which is organized in various subsequent stages. Every stage contains some message for the compositional algorithms and some directions for expressive performance. An example of notes employed for recalling the score requirements to the user is shown at Figure 6.6, where pitch events (*la lungo*, long A), dynamics (p, mp, etc.), ring modulation effects (MP, P, etc.), frequency shift ($\wedge$) and second frequency shift ($+$) are noted in summary form. As far as concerns the interactive score production, the user is free to decide when to trigger to cue button and hence to progress to the next score stage, or when to stop the performance. The expressive performance section is subdivided in the score alignment phase, where the user changes the active region following the score directions, and in the expressive alignment phase, where the user, once has reached the active zone, moves inside it searching the best musical result. In this case the musical gesture is dependent on the expressive intentions of the user and acts both as the cause and the effect of sound production (Cadoz et al., 2000). In this way action and perception are linked in a continuous loop that drives the user's movements and timings.

## 6.2 Case Study 5: Hand Composer, Gesture-driven Composition Machines

Hand Composer is a gesture-driven composition system, based on the analysis of the existing relationships between music generative models and musical composition in the context of the historical background of the $20^{th}$ century music. The system framework is based on a number of interactive machines performing various patterns of music composition, and producing a stream of MIDI data to be used by a Disklavier.[7] Hand gestural input, captured by the Leap Motion sensor,[8] can control some parameters of the music composing machines, changing in real time their musical output.[9] Hand Composer reinterprets the same goals of Disembodied Voices with some important changes:

1. It employs a three-dimensional sensor-centered space. Action happens in a limited space positioned above the sensor and not in the user's peripersonal space as in Disembodied Voices.

2. It expands the concept of interactive composition in that the interaction is not limited to the score's stages duration but influences the composition's structure. At the same time, the expressive interaction is much more limited than in Disembodied Voices because it is only limited to control the output's dynamic level.

### 6.2.1 How it Works

Hand Composer allows interactive composition and musical expressive interaction. To reach these goals it employs two basic gestures: the continuous palm shift, obtained by moving the hand inside the sensor's active three-dimensional region (Line), and the palm $z$ plane threshold crossing, obtained by moving swiftly the hand towards the computer screen (Point). These gestures are interpreted by higher level composition machines in various ways, according to the directions of a previously arranged musical score, which is composed of different subsequent

[7]The Disklavier is a standard acoustic piano, except that it can also employ electromechanical solenoids to move key and pedals independently of any human performer. Thus the Disklavier can be played in the traditional way, but can also be controlled by MIDI messages sent from a computer through a USB cable or stored in memory units or CD's.

[8]The Leap Motion controller is a small USB peripheral which is placed in front of the laptop. The device scans a region in the shape of an inverted pyramid centered at the device's middle point and expanding upwards for about 60 cm (2 feet).

[9]A musical performance of Hand Composer may be found at `https://youtu.be/mdsn9_5Ig_A`

stages. The user employs the right hand for sensor interaction and the left hand for computer keyboard input aimed at stage progress.

### 6.2.2 Related Work

Gesture-driven composing systems that produce symbolic output are not very common in literature. Many of them present strong stylistic limitations due to a general lack of music theory background. Also the relationships between gesture and musical meaning seem often to be rather complimentary. Cyber Composer (Ip et al., 2005) for instance is devoted uniquely to tonal music production with gloved hand motion controlling pitch, volume and other parameters. The abstract nature of musical elements (like tonality or cadence selection) doesn't allow any tight relationship between gesture and musical information, making gestural interaction rather unnatural. The Bigbang Rubette (Thalmann and Mazzola, 2008) is the gestural extension of the Rubato Composer software, which relies on five standard transformation types: two dimensional translations, rotations, shearing, dilation and reflection. The software allows all these operations using gestural mouse commands or common two-dimensional surfaces finger gestures. As no musical insight is provided by the authors, the above mentioned software can be considered more a music editor than a musical composition software. Kristoffer Jensen in (Jensen, 2011) presents some interesting ideas about temporal perception and pitch gestures generative models, but his algorithm is uniquely devoted to melodic contour reconstruction. The Crosshole (Şentürk et al., 2012) is a Kinect operated meta-instrument which shows three levels of gestural interactivity: the chord structure, the arpeggio control and the timbral manipulation. In the Harmonic Navigator (Manaris et al., 2013) gestures are employed to display, select and stop playing chords from a visual interface fed by a genetic algorithm. Again, in these last two examples, the authors deal with very abstract concepts which are difficult to be mapped in a gestural meaningful way. At the end, the recent Leap Motion Muse[10] is a composition system designed for the Leap controller. The user is presented a matrix of sixteen buttons to be selected by finger motion. A hand gesture rotates the cubic matrix, allowing further selection possibilities. The produced result seems to be nothing more than a mix of pre-defined musical output, which can be recorded, posted and shared on social networks.

### 6.2.3 Hand Composer Musical Information

The framework of Hand Composer is based on the definition of a number of music composition models taken from the $20^{th}$ century music production. Music features that can be described as global music situations in a clear way and that can be outlined as the fingerprint of an author's musical expression are considered. A musical fingerprint can be defined as a perceivable, repeated and coherent musical behavior, made of various musical features which occur for a certain time and which are often found in an author or in a specified period of author's production.

[10]Leap Motion Muse `http://boulangerlabs.com/wp/product/muse`

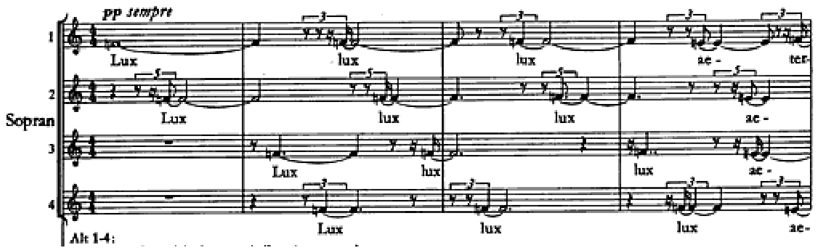



**Figure 6.7:** *An excerpt from György Ligeti's Lux Aeterna (1968), showing an example of micropoliphony technique.*

#### 6.2.3.1 The Ligeti model: adjustable pitch and.

The Ligeti[11] model refers to progressively enlarging bands of sounds whose articulations, density and internal speed depend on the instruments employed. Among author's large production it is possible to find various examples of this feature in a version for vocal ensemble (Lux Aeterna, 1968, Figure 6.7), chamber instruments (Kammerkonzert, 1970) and harpsichord (Continuum, 1968). The name of such peculiar musical situation is micropoliphony, which emphasizes the initially very limited range of melodic movements of every band component. The various parts have a completely autonomous attack times and event's length. As time goes by, little progressive range shifts may lead to a continuous widening of the band, including more and more new pitches.[12]

#### 6.2.3.2 The Webern model: single pitches with melodic leaps.

The Webern[13] model is inspired by well-known very frequent fingerprint musical situations where single pitches appear at different instrumental registers. The Webern's fingerprint is characterized by the intertwined horizontal display of various four-notes microcells arranged at melodic intervals seldom greater than the octave. Pitchs and durations, which are strictly dependent from the musical row employed by the author, give life to the so called Webern's constellations, named this way to express the sparseness and the wide space extension of pitch's appearance, as can be seen in Figure 6.8.

### 6.2.4 The two Models Geometrical Interpretation

In the algorithmic models the musical information is reinterpreted and conveniently adapted. The Ligeti model disregards inner instrumental articulations, but rather concentrates on the band

[11] György Ligeti (1923 - 2006) is an Hungarian composer famous for his composition technique named micropoliphony and for renowned works like Atmosphères (1967) and Lux Aeterna (1968).

[12] Some audio example of the Ligeti and Webern models is available at `http://www.dei.unipd.it/~mandanici/handcomposer.html`.

[13] Anton Webern (1883-1945) was an Austrian composer who studied composition with Arnold Schönberg from whom he learned the serial composition technique. He interpreted serialism in the most extreme way, opening the way to the so-called integral serialism that characterized an important part of the musical production in Europe after World War II.

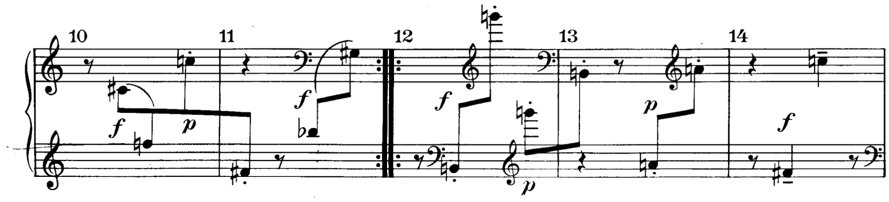



**Figure 6.8:** *An excerpt from Anton Webern's Variationen for piano op. 27 (1936). This example shows the typical wide range melodic display of Webern's constellations.*

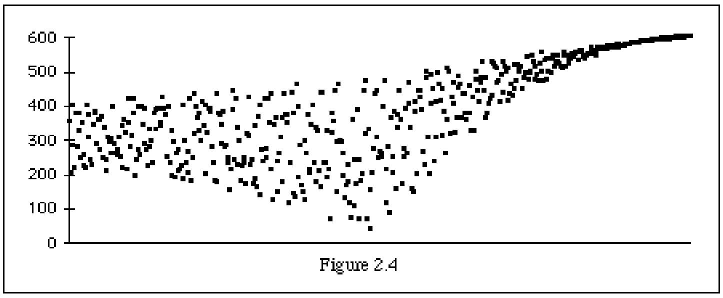



**Figure 6.9:** *An example of tendency mask for shaped composition of random events. Source* `http://www2.ak.tu-berlin.de/~abartetzki`

production as a global phenomenon. The Ligeti's model consists of a band of pitches with various starting points, ranges and densities. The band is implemented by a very fast production of single sounds with adjustable bandwidth constraints, starting from a minimum (a single pitch) to a maximum customizable range. The algorithmic model of a shaped composition of random events has been named tendency mask technique and has been theorized by G.M. Koenig in 1966.[14] It is defined by two envelopes for lower and higher boundaries, with at least another two controls for values distribution and density (probability tables, random walks, etc.). The geometrical interpretation of a band of pitches may look like the one depicted in Figure 6.9. Similarly, in the Webern's model the extreme determinacy of pitches and durations typical of the serial composition is not considered. Rather, the author's fingerprint is interpreted as a random choice of the $x, y$ couple (representing midi pitches and velocities) in a custom field of values. Also the event's duration is randomized in a range from 50 to 5000 ms.

### 6.2.5 Hand Composer Interface

The Hand Composer interactive space has the form of an inverted pyramid centered at the Leap Motion's middle point. The $x$ axis, in red in Figure 6.10, corresponds to midi pitches and the $y$ axis (in green) to velocities. A hand waving inside the active region produces a Cartesian couple of coordinates of the palm position at such a definite degree as to be able to select along the $x$ axis a single pitch in a range of midi notes from 30 to 100. For each $x$ position it is possible

[14]G.M. Koenig `http://www.koenigproject.nl/indexe.htm`

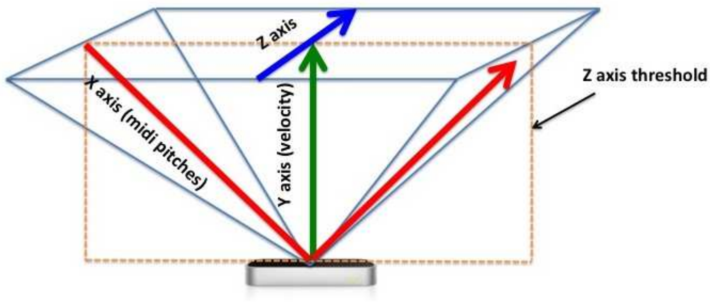



**Figure 6.10:** *The three-dimensional active region of the Leap Motion sensor with Hand Composer data mapping. The inverted pyramid centered at the device's middle point is subdivided by a $z$ axis threshold that acts as trigger point.*

to choose a $y$ position, moving the hand up or down a vertical axis, producing in this way the midi pitch's velocity. This is the first interaction modality that produces a continuos data flow of the waving hand (Line) and is used inside the Ligeti music composition machine. By waving the hand gently upward and downward a maximum and a minimum range for the pitch band is generated. The gesture conveys also other important information like the trend of band growth and speed of changes. It is to be observed that the hand interaction in this case may be much richer in available information than a simple automated production, which, to obtain the same results, would need a greater amount of data. The second interaction modality is a trigger that is produced by assigning a threshold to the $z$ axis. Every time the palm overcomes the threshold value, a trigger is produced together with the coordinates of the palm at the triggering moment (Point). This gesture can be used by the Webern model to produce couples of pitch/velocity as data input for pitches constellation.

### 6.2.6 Hand Composer System Architecture

The Hand Composer system consists of a Leap Motion sensor for hand motion capture, a Max/MSP patch for motion data processing and MIDI data generation, and of a Disklavier for music production. The core of the Hand Composer system are the music composition machines, which

| *PARAMETERS* | **LIGETI MACHINE** | **WEBERN MACHINE** |
|---|---|---|
| *Speed* | message | message/Point |
| *Start pitch* | message/Point | message/Point |
| *Velocity* | message/Line | message/Point |
| *Duration* | message | message |
| *Band control* | message/Peak | |

**Table 6.3:** *Table of the musical parameters employed by the Ligeti and Webern machines and their input modalities.*

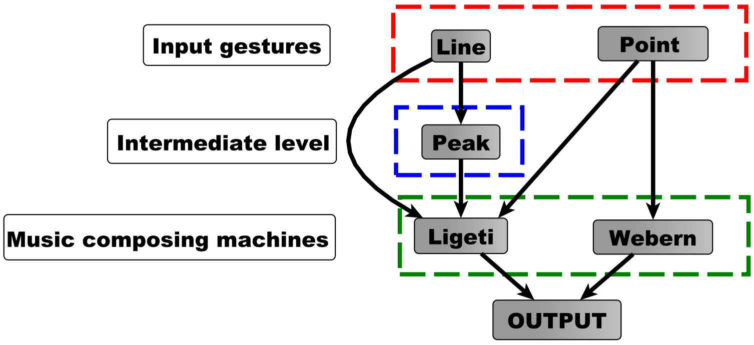



**Figure 6.11:** *The three levels of the Hand Composer music production system. Input gestures are employed to feed the music composing machines or directly or passing through an intermediate level for further processing.*

are high-level producing tools defined by a number of musical parameters. The parameters may be determined by a programmed message or may be produced in real-time by hands movement. As there are many possibilities to influence the musical output by gestural interaction - but they can only be activated one at a time - a decision about how to interpret gestural data is needed. This is made through the definition of a number of stages annotated in a musical score, whose progress is operated by the user's left hand interaction with an appointed computer key. Thus, for every stage the user's knows how her/his gestures will be interpreted by the system and can so properly adapt her/his behavior. The music composing machine system is organized in three levels, as shown in Figure 6.11. The first level corresponds to the two basic input modalities, Point and Line, described in Section 6.2.5. Line produces a continuous stream of couple of $x, y$ data, whereas Point outputs the same couple only when the $z$ threshold is surpassed. The intermediate level builds upon gestural basic input and processes it to obtain useful data for the composing machines. As shown in Figure 6.11, in the present implementation it is represented by Peak, which looks if a line is ascending or descending and computes the difference between the two peaks. The musical parameters necessary to the music composing machines are shown in Table 6.3. *Speed* controls the band's density, that is the number of pitches produced per minute by the Ligeti machine, whereas in the Webern machine the event's speed may be ruled by a randomic time message or by user's interaction. *Start pitch* controls the beginning point of the pitch band in the Ligeti machine or the pitch height of one of Webern's sparse constellation's events. *Velocity* controls the general events dynamics and can be ruled by a message or by hand's vertical movements on the $y$ axis (Line for the Ligeti machine and Point for the Webern machine). *Duration* is controlled by a message. *Band control* rules the band's trend (how much and how fast it grows or shrinks) and may be fed by a series of messages or by hand interaction through Peak.

### 6.2.7 Hand Composer Application Framework

The Hand Composer application framework is depicted in Table 6.4. As in Disembodied Voices, the Hand Composer's user acts both as a composer and as a performer, though with more possi-

| HAND COMPOSER APPLICATION FRAMEWORK | | | |
|---|---|---|---|
| **Aim** | | **Interactive algorithmic composition** | **Expressive performance** |
| **Interface** | *Musical Information* | 1) The Ligeti model<br>2) The Webern model | The musical conductor model |
| | *Geometric Interpretation* | 1) Scattered points in the interaction space<br>2) Tendency mask of random events | A vertical line |
| | *Spatial Positioning* | 1) $xy$ data of the whole active region<br>2) The $z$ axis threshold | Along the vertical axis starting from the centre of the sensor to the limit of the active region |
| **Interaction** | *Where to move* | To the position that better fits the structural changes of the composition | To the position that better fits the dynamics of the event |
| | *When to move* | When new event changes need to be generated | Whenever an expressive change is needed |

**Table 6.4:** *The table shows the Hand Composer aim, musical information, geometrical interpretation and spatial positioning of active regions and thresholds for both the algorithmic composition and expressive performance interaction. The interaction paradigm is completely dependent on the score's directions.*

bilities for the benefit of the interactive composition at the expense of the expressive interaction. In fact, contrary to Disembodied Voices, Hand Composer's composing machines are open to user interaction. Thus, the whole application framework has been developed both for interactive algorithmic compositions and expressive performance. Contrary to Disembodied Voices, the two functions cannot be performed in the same time. The routing of the hand's motion data is decided stage by stage depending on the directions of the musical score. Thus, user's movements are strictly subjected to the aim assigned by the score.

# Chapter 7

# The Harmonic Walk Assessment

Many questions arise when employing an interactive environment like Harmonic Walk for learning purposes. The main of them concern the learning power of bodily interaction. Embodied cognition is grounded on the idea that a great part of human knowledge is based upon information originating from bodily interaction with the physical world, and motion-based application design relies precisely on this. Then: how does bodily interaction relate to abstract knowledge? Is it possible to make embodied knowledge become explicit? Is enactive learning more effective than explicit information? To investigate these questions a two session experimental test approach was planned and various observations about user's behavior were collected and analyzed. The results are showcased and discussed in the following Sections.

## 7.1 Learning in the Harmonic Walk Environment (First session)

Harmonic Walk is a music learning application that can be employed both by professional musicians and by people with no particular musical training. In both cases the user leverages on her/his implicit knowledge of harmonic language and musical structure to understand how the application works and to accomplish the required tasks. Whatever the level of musical knowledge of the user, s/he can discover, verify and confirm her/his musical knowledge during the experience. The Harmonic Walk learning test was organized at the Catholic Institute Barbarigo in Padova (Italy) in December 2014 with the aim to gauge at what extent the application's design is successful in introducing the user to the tonal music features without any previous information about the application's task and materials. To do so, the subjects' success in the melody segmentation recognition and in the melody harmonization tasks was measured. Moreover, to evaluate how much of the manifold musical contents of the Harmonic Walk had been perceived by the students, a questionnaire for both quantitative and qualitative assessments was fulfilled by the participants. In particular, the test aimed at verifying if the subjects could gain some musical knowledge and consciousness about the musical features employed after the experience in the Harmonic Walk environment.

### 7.1.1 Subjects

A total number of twenty-two high school students between sixteen and twenty years old took part in the test. The students were equally subdivided into two different groups: one musically trained and the other not. The first group was taken from classes belonging to the music high school, with specific instrumental and music theory programs. The second was chosen from classes belonging to various kinds of high schools with no music programs in their curricula. Among these students with no musical practice at all were asked to take part in the test. To preserve clarity, only subjects with a declared state of disability were excluded from the test.

### 7.1.2 Materials

The choice of the music genre and of the music piece which could fit better our test was not trivial. Pop music was chosen for its popularity and because the use of the sung words can help learning, memorization, localization of melodic qualities and performance. Actual hits were bypassed for at least two reasons: This type of songs could be too restricted for the test's needs (i.e., it could not offer enough options regarding chords complexity, harmonic rhythm, and so on); moreover, as not all young people listen to the same kind of music, the choice could not guarantee that all the students know the same song. Thus the search of the most suitable song started with the definition of its ideal features:

a) the chords employed must belong all to the same key, to fit the six roots harmonic space explained in Section 5.1.5;

b) the song's harmonic progression should include, in addition to two or three primary chords, at least one - but no more than two - parallel chords, to offer a medium level of harmonic complexity;

c) the chord's change (i.e., the harmonic rhythm) must be not too fast, in order to allow easy body movements in the physical space;

d) the song must be very popular, clear and easy in the melodic structure.

After an accurate search, Adriano Celentano's "Il ragazzo della via Gluck"[1] was chosen. The song's musical structure is summarized in Table 7.1. The first musical phrase of the song is composed of eleven segments built upon three chords (two primary and one parallel). The column of the "Syllables of change" shows the syllables of the text corresponding to the harmonic changes, i.e., the points where the melodic segmentation occurs. The table shows that the segmentation points are ten, as well as the number of chord changes. The column "Duration" outlines the number of bars occupied by the different song's units. As can be seen, the harmonic rhythm is not constant, i.e., the song's unit have not all the same duration.

[1]Adriano Celentano (1938) is a popular Italian singer, very known also abroad. He has begun his career in the sixties. Il ragazzo della via Gluck, written in 1966, is one of his greatest hits.

| N. of Seg. | Syll. of change | Chord | Duration |
|---|---|---|---|
| 1 | - | I | 1 |
| 2 | Gluck | V | 2 |
| 3 | me | I | 2 |
| 4 | ttà | V | 2 |
| 5 | ge | I | 2 |
| 6 | ten | V | 2 |
| 7 | ttà | I | 1 |
| 8 | vrai | VI | 3 |
| 9 | qui | I | 3 |
| 10 | var | VI | 3 |
| 11 | ti | I | 3 |

**Table 7.1:** *Table of musical segments, syllables of change, chords and harmonic rhythm of Adriano Celentano's Il ragazzo della via Gluck.*

### 7.1.3 Scenario

In Section 5.1.5 two musical features, melodic segmentation and the tonal harmonic space have been discussed, with the two related geometrical interpretations and masks (the straight line and the circular ring). For the Harmonic Walk's learning test a three phase approach to the tonal melody accompaniment has been planned, as shown in Figure 7.1. In the first phase the user is presented with the straight path along the borders of the active area, corresponding to the source-path-goal schema shown in Figure 5.5. Every step through the interactive landmarks produces the sound of the portion of the audio file corresponding to the segmentation described in Section 5.1.4. If the user fails to step to the next zone in time (i.e. if s/he is not aware of the song's harmonic rhythm), the audio file ends and the performance is interrupted. In the second phase the user enters the circular ring containing the six roots discussed in Section 5.1.5, with the aim to explore the harmonic space and to match the spatial position to the chords sound and function. In the last phase, starting from the melodic segmentation knowledge acquired in the first phase and from the experience of the harmonic space acquired in the second phase, the user can now try to find the chords useful for the melody accompaniment of the composition. As usually there is not a unique solution to the melody harmonization, the user can experiment various chord sequences and choose the one s/he prefers. The task is considered complete and successful when the user harmonizes the whole song melodic excerpt without interruptions and in a way s/he considers valid.

### 7.1.4 Method

The high school students were tested individually in private sessions where only the test conductor and the music teacher were present. No previous information about the song used during the test was provided to the subjects. Each student was presented with some written instructions

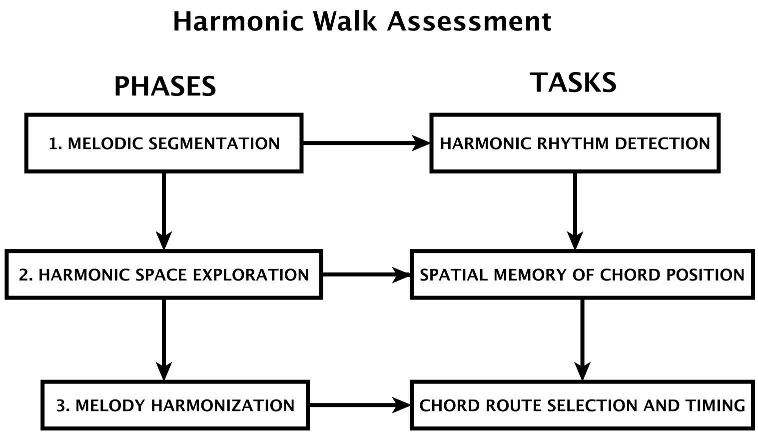



**Figure 7.1:** *Harmonic Walk assessment phases and tasks.*

about the tasks s/he had to accomplish, whereas a short demonstration about the environment and its interaction modalities was provided by the test conductor. After this, the student undergoes the first phase of the test, during which s/he listens to the song, trying to understand where the units sectioning points are and trying to link them through an appropriate pace timing. The student is given a maximum of five minutes to conclude the first phase of the test. In the second phase the student enters the circle of the song's harmonic space and explores it freely. S/he has three minutes to listen to the chords and to remember their position in space. In the third phase the student is given the text of the song and, with the help of the teacher who sings the melody, tries to find the chords which can fit the song. As soon as s/he feels that the chord is not correct, the song begins again and continues till the end or till the next mistake. The student is given five minutes to complete this task. After time expires or the last task is accomplished, the student is presented to a questionnaire for both quantitative and qualitative feedback about the test. The test conductor fulfills also a form for qualitative assessment and further observations. Moreover, for each of the three phases of the test, the system records the hit zones and the overall time employed by the subject.

### 7.1.5 The Questionnaire

The questionnaire prepared for the test subjects contains open ended interviews, Likert-scale surveys[2] and subjects opinions. It focuses on the first and third phase of the test and it is designed to collect both quantitative and qualitative data from the test subjects. For the first phase an interview about the detection of the song's syllables has been prepared. No inquiry about the second phase of the test has been made because data were collected through direct observation and subjects' movement logs. For the third phase three open-ended interviews were proposed: one about how many times the subject did a chord change; the second about how many overall

[2]The Likert-scale is a technique used to scale responses in survey research. It employs at least five items ranging from negative to positive statements, allowing so an easy and fast computation of the test's results.

| | Quantitative Assessment | | | Qualitative Assessment |
|---|---|---|---|---|
| | Tests Evaluations | Interview | Likert-scale | Subjects Opinion |
| $1^{st}$ **phase** | - | Syllables of change | - | - |
| $2^{nd}$ **phase** | 1) Direct observations<br>2) Subjects movement logs | - | - | - |
| $3^{rd}$ **phase** | 1) Number of partial harmonizations (PH)<br>2) Number of complete harmonizations (CH)<br>3) Number of failed harmonizations (NH) | 1) Numbers of chord changes<br>2) Total number of chord employed<br>3) Drawing of the harmonizing route | - | - |
| **General evaluation** | - | - | 1) Pleasantness<br>2) Utility | 1) General opinion about the application<br>2) Principles upon which the application is based |

**Table 7.2:** *General table of assessment tools utilized in the Harmonic Walk learning test. The tools are tests evaluations, open-ended interviews, Likert-scale reviews and opinion questions.*

chords s/he used; in the third the participant was asked to draw on a plotted schema of the six chord circular ring the route s/he did for harmonizing the melody. At last, to collect some general user evaluation about the Harmonic Walk, two Likert-scale surveys about the test experience pleasantness and utility were proposed. Both surveys have five items. From "Very disagreeable" to "Very pleasant" for the first; from "Incomprehensible" to "Very useful" for the second. As this is the first rigorous assessment test for Harmonic Walk, the item's definition could not rely upon previously collected users' statements. Thus neutral items were employed. The questionnaire ends with two opinion questions about the subjects' overall feeling and about the musical principles employed in the test.

### 7.1.6 Harmonic Walk Learning Test Results

In this Subsection the test results are presented. They are organized as shown in the overall scheme of Table 7.2.

#### 7.1.6.1 First phase: syllables of change detection.

The harmonic changes perception is the first cognitive result aspected after the test proposed in the first phase. The user had to wait the end of the song's harmonic unit before linking the next with her/his step. In this way s/he should have reinforced his/her perception of the harmonic changes. To verify if this knowledge has actually been realized by the test participants, they were asked to underline the song syllables where they felt a harmony change occurred. The means of the hits are shown in Figure 7.2. As expected, the results of Musicians are better than those of Not Musicians. The means change significantly also depending on the song's harmonic unit. As a matter of fact the hits' means drop dramatically among Not Musicians (but also, less dramatically, among Musicians) as soon as the song enters the last four harmonic changes, those involving the parallel harmony (vi degree), whereas it is rather high at the beginning of the song.

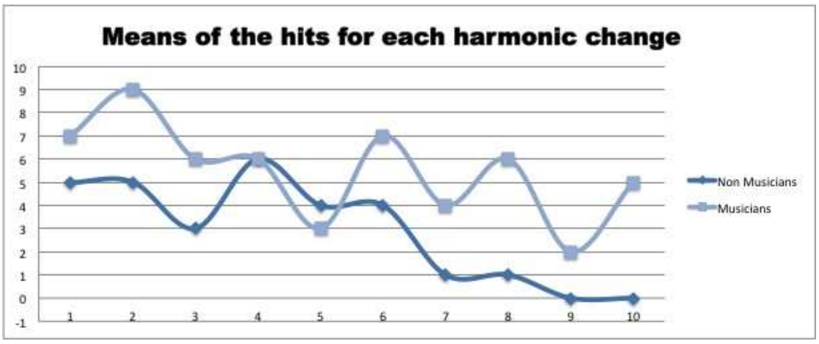



**Figure 7.2:** *Means of the hits over ten harmonic changes obtained by Not Musicians (blue line) and Musicians (azure line).*

#### 7.1.6.2 Second phase: harmonic space exploration.

In the phase of the harmonic space exploration the subject has the possibility to explore the circular ring where the chords are laid. This experience is crucial for the success of the next required task - the melody harmonization - because it is here where the subject builds her/his spatial map of the harmonic relationships. Therefore, it is important to analyze what kind of movements subjects did while exploring the circular ring. From direct observation and from the analysis of movements recorded by the system during the test, two different approaches to the circular ring exploration were found. In the first approach the subject follows the shape of the circle starting from the left or from the right side. S/he touches all the six zones in sequence, thus obtaining the harmonic progressions shown below:

I - V - III - VI - II - IV
or
I - IV - II - VI - III - V

These are very long progressions without the repetition of the tonic chord, which is a fundamental perceptual anchorage point for tonal progressions. If it fails to appear for a certain time, the harmonic progressions lose their direction. Simply walking around the circle does not reinforce tonal chords localization as expected. On the contrary, the second approach emphasizes movements where the frequent return to the tonic chord is present. This exploration approach is nearer to the actual tonal composition's structure, where the tonic frequently occurs and where chordal progressions are often repeated. Returning on the tonic chord helps to distinguish harmonic functions; furthermore, the repetition of the movement among two well identified space regions trains the body to perform simple and strong harmonic progressions. As demonstrated by the recorded paths (see some examples in Figure 7.3) and by direct observation of the subjects' movements, the exploratory phase has been only seldom characterized by recursive movements on the tonic chord. Sometimes movements like 1 - 6 - 1 (tonic, dominant, tonic) or 1 - 2 - 1 (tonic, subdominant, tonic) were registered, but never movements like 1 - 4 - 1 (tonic, minor parallel, tonic) or 1 - 5 - 1 (tonic, mediant, tonic), nor perceptually strongest progressions like 1-2-6-1 (tonic, subdominant, dominant, tonic).

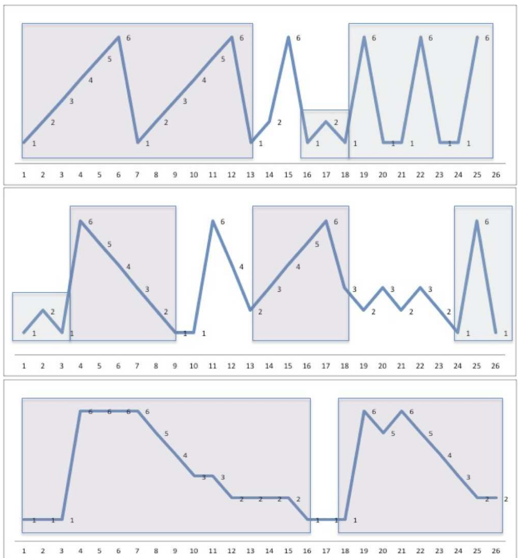


**Figure 7.3:** *Some exploratory paths recorded during the second phase of the Harmonic Walk learning tests. The azure areas correspond to the recursive movements, 1-6-1 or 1-2-1, whereas the violet areas correspond to the circular paths.*

#### 7.1.6.3 Third phase: partial or complete harmonizations.

The melody harmonization is the conclusive main task of the third phase of the application's test. The test conductor reports a complete successful harmonization when the subject can find the chord sequence that s/he considers a good harmonic solution for the song and that s/he is able to repeat (CH). When the subject succeeds in harmonizing only the first part of the melody, where only tonic and dominant chords are employed, a partial harmonization is reported (PH). If the subject, also after many trials, cannot provide an ultimate harmonized version, a failure is reported (NH). None of the complete or partial successful harmonizations differs from the model proposed in the original song. The full harmonization task was accomplished by one Not-Musician and by one Musician subject. Other four musician subjects accomplished the task only in the first part of the song, that involved only the tonic and the dominant chord. Thus a complete or partially successful accomplishment task was recorded for six subjects, that is the 27% of the total. Again, the results show the difficulty in the perception of the minor parallel harmonic areas. As can be seen in Table 7.3, among the six harmonizations only two were complete, while the other four could accompany the song only in its simplest part, the one involving the tonic and dominant chords.

#### 7.1.6.4 Third phase: chord changes and number of chords employed.

To find out if the tested subjects have gained some knowledge during the test, some questions about how many times the subject changed the chord and how many different chords s/he employed during the third phase of the test (the melody harmonization) were posed. The results are shown in the columns "Changes" and "'Emp. Ch." (number of employed chords) in Table 7.4.

| Harmonization | Not Musicians | Musicians |
|---|---|---|
| Complete Harmonization (CH) | 1 | 1 |
| Partial Harmonization (PH) | 0 | 4 |
| No Harmonization (NH) | 10 | 6 |
| Total CH+PH | 1 (4,5%) | 5 (22,7%) |

**Table 7.3:** *Table of complete, partially successful and unsuccessful harmonizations over twenty-two subjects. In the last line the total of complete and partial harmonizations is reported.*

The number of chord changes identified by the subjects hits a mean of 4.8 for Not Musicians and of 7 for Musicians upon a total of ten changes. The number of chords employed identified by the subjects hits a mean of 3.5 for Not Musicians and of 4.2 for Musicians upon a total of three employed chords. The ability to detect the number of chord changes as different from the overall number of chords employed is crucial for the tonal language musical knowledge, because shows the difference between the chord changes (ten in eleven harmonic units in the straight path) and the number of places to reach in the circular ring (three, corresponding to the tonic, dominant and minor parallel chord). This difference is exactly the effect of the recurrence of the tonic-dominant and tonic-minor parallel progressions. In general, a scarce correlation between the partial or total success of the subjects in the harmonization task, and the correctness of their answers in the questionnaire has been observed, as can be seen in Table 7.5, where the stars outline the correct answers.

#### 7.1.6.5 Third phase: the harmonization route.

In a third interview the subjects were asked to draw the path they performed during the harmonization test on the circular ring, marking it with arrows or numbers. The records are divided in groups depending on the number of right chord sequences identified by the subjects, as shown in Table 7.6. In the first row the number of not answered questions is reported. In the second row the number of answers where only the first chord was right and the second was wrong. This means that the subjects failed to identify even the first chord progression of the song, which was instead identified by the subjects of the third row. The fourth row's subjects show to have understood that the sequence returned on the first degree at least once, while the fifths row's subjects caught that the sequence was repeated more than once. At least, identifying more than six right chords means that the subject has found also the vi degree progression. Nonetheless, there is never a full correspondence between the right or partially right melody harmonization subject and the identification of the right sequence, as can be seen in Table 7.7. Comparing the sequences reported in Table 7.7 with the correct chord sequence

1 6 1 6 1 6 1 4 1 4 1

it can be seen that in all the sequences there is always some mistake, or some missing chord (like in subject 4 where there are only the last two chords missing).

| Not Musicians | Syllables (10) | Changes (10) | Emp. Ch. (3) |
|---|---|---|---|
| 1 | 6 | 7 | 5 |
| 2 | - | - | - |
| 3 | 6 | 4 | 2 |
| 4* | 2 | - | 4 |
| 5 | - | - | - |
| 6 | 0 | - | 6 |
| 7 | 5 | - | 4 |
| 8 | 7 | 4 | 2 |
| 9 | 0 | - | - |
| 10 | 3 | 6 | 2 |
| 11 | - | 3 | 3 |
| **Mean** | 3.6 | 4.8 | 3.5 |
| **Correct Answers** | 0 (11) | 0 (11) | 1 (11) |
| **Musicians** | **Syllables (10)** | **Changes (10)** | **Emp. Ch. (3)** |
| 1** | - | 9 | 4 |
| 2** | 8 | 10 | 12 |
| 3 | 6 | 7 | 3 |
| 4 | 4 | 5 | 4 |
| 5** | 7 | 9 | 3 |
| 6* | 4 | 7 | 3 |
| 7 | 10 | 10 | 5 |
| 8 | 7 | 10 | 4 |
| 9** | 4 | 4 | 2 |
| 10 | 3 | 4 | 4 |
| 11 | 2 | - | 3 |
| **Mean** | 5.5 | 7 | 4.2 |
| **Correct Answers** | 1 (11) | 3 (11) | 4 (11) |

**Table 7.4:** *Table of the number of syllables of harmonic changes, the number of chord changes and the number of employed chords recorded by Not Musician and Musician subjects, with the correct answer in brackets. The subjects with the asterisk are those completely successful in the harmonization task, whereas the subjects with the double asterisk are those partially successful in the harmonization task. The mean and the number of the correct answers upon the number of subjects is shown in the last lines of the table.*

| Subject | Harm. | Syll. | Changes | Emp. Ch. |
|---|---|---|---|---|
| No. 4 (NM) | CH | 2 | - | 4 |
| No. 1 (M) | PH | - | 9 | 4 |
| No. 2 (M) | PH | 8 | 10* | 12 |
| No. 4 (M) | PH | 7 | 9 | 3* |
| No. 6 (M) | CH | 4 | 7 | 3* |
| No. 9 (M) | PH | 4 | 4 | 2 |

**Table 7.5:** *Table of answers to questions about the syllables of change, the number of harmonic changes and the number of employed chords obtained by partially (PH) or complete (CH) successful subjects in the harmonization task.*

| Sequences | Not Musicians | Musicians |
|---|---|---|
| No answer | 1 (9%) | 0 |
| No Sequence | 3 (27%) | 0 |
| First 2 Chords | 5 (45%) | 5 (45%) |
| First 3 Chords | 2 (18%) | 0 |
| More than 3 Chords | 0 | 2 (18%) |
| More than 6 Chords | 0 | 4 (36%) |

**Table 7.6:** *Table of chord sequences identified by the subjects during the harmonization task.*

#### 7.1.6.6 General evaluation: Harmonic Walk pleasantness and utility.

Two Likert-scale surveys were included in the questionnaire. The first survey aimed at recording a judgement about the application's pleasantness; the second survey was about the utility of the application. The results are respectively shown in Figure 7.4. The mean of ratings for application's pleasantness is 3.4 for both Not Musicians and Musicians. The mean for the rating for the application's utility is 2.9 for Not Musicians and 3.6 for Musicians, showing that for both reviews subjects expressed a neutral opinion.

#### 7.1.6.7 General evaluation: qualitative assessment.

Data about qualitative assessment come from two opinion questions contained in the questionnaire. The first question aimed at recording the subject's perception of the application. Some subjects judgements, ideas, impressions and feelings about the application and its use are partially reported in Table 7.8. Other judgements, expressed only by one subject, were: strange, incomprehensible, frustrating, complicated, indifferent, engaging, nice and beautiful. It is also interesting to look at the reasons why such judgements were expressed. Harmonic Walk was considered useful by Not-Musicians to develop the listening memory, to evaluate the understanding of musical features, to understand the musical experience, to keep in time while playing with

| Subject | CH or PH | Recorded Sequence |
|---|---|---|
| No. 4 (NM) | CH | 1-(2-1-2-4-1) |
| No. 1 (M) | PH | 1-6-1-6-1-6-(4 -3-1) |
| No. 2 (M) | PH | 1-6-1-6-1-(4-3-4-3-5-6-1) |
| No. 4 (M) | PH | 1-6-1-6-1-6-1-4-1 .... |
| No. 6 (M) | CH | 1-6-1-6-1-6-1-(6-1-4-1-4) |
| No. 9 (M) | PH | 1-6 ... |

**Table 7.7:** *Table of chord sequences identified by the complete and partial successful subjects during the harmonization task. In brackets are reported the wrong parts of the sequence.*

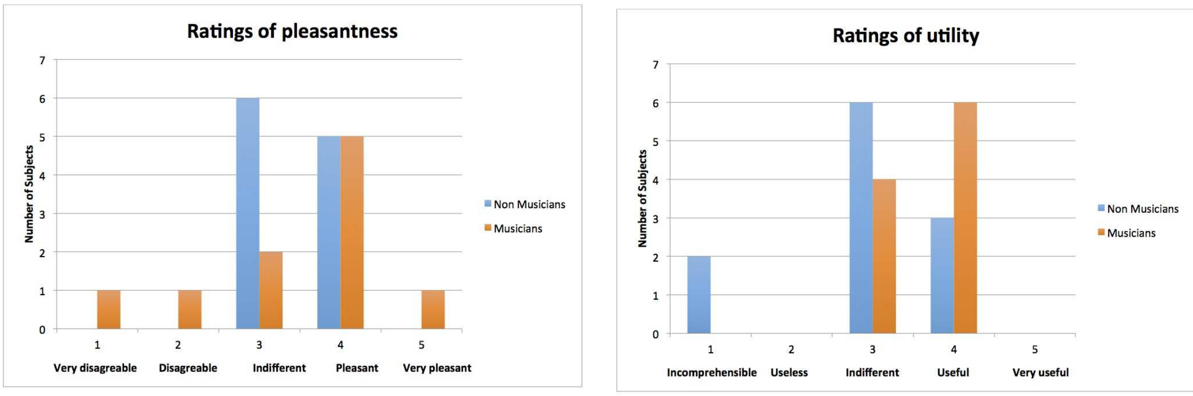



**Figure 7.4:** *Chart of ratings of application's pleasantness and utility.*

others, to work with children, to keep coordination and precision in music performance, to learn tonal harmony in practice, to develop aural skills, to test your own musical skills. It was considered interesting by Not Musicians to discover how complex may be a popular song and to see how the body reacts to music. It was also considered beautiful by one musician subject because it can introduce people to music. As can be seen, subjects were highly impressed by the application and spontaneously reported a great number of ideas about the experience during the test. The second question aimed at discovering what musical principles subjects thought were involved in the test. As can be seen in Table 7.9, musical memory and rhythmic skills are the strongest perceived musical features employed in the design of the Harmonic Walk.

### 7.1.7 First Session Test Results Discussion

To make a rigorous test of the Harmonic Walk application very hard and strict assessment conditions were established:

a) high level of musical features involved in the application design;

| **Judgements** | **Not Musicians** | **Musicians** |
|---|---|---|
| Useful | 3 | 4 |
| Interesting | 3 | 1 |
| Difficult | 2 | 1 |
| Amusing | 1 | 2 |
| Useless | 2 | 0 |

**Table 7.8:** *Table of the subjects' ideas about the application.*

| **Principles** | **Not Musicians** | **Musicians** |
|---|---|---|
| Musical memory | 4 | 2 |
| Rhythmic skills | 1 | 3 |
| Body reaction to music | 2 | 1 |
| Song segmentation | 2 | 1 |
| Aural skills | 1 | 1 |
| Connections between melody and chords | 0 | 1 |

**Table 7.9:** *Table of perceived principles upon which the application was designed.*

b) no previous hints about the application cognitive contents, nor about the song employed during the test;

c) strict test protocol and time constraints.

Of course this choice greatly affected the chance of success of the subjects, who could rely solely upon their previously acquired knowledge of the musical structural elements involved in the test. Thus, the test does not bring great success expectation, but rather suggests what is the learning potential of a system like Harmonic Walk. The general results of the $1^{st}$ session of Harmonic Walk learning test, summarized in Table 7.10, are the following:

1. *The recognition of the syllables of changes* hits a mean of 3.6 for Not Musicians and of 5.5 for Musicians, over ten syllables. The mean is rather low, but, as shown in Figure 7.2, the hits are higher in the first part of the songs and decrease depending on the song's harmonic area (minor parallel area), demonstrating a difficulty in parallel harmonies perception. Nonetheless, as this area covers the second part of the song excerpt, the reason of the drop of the number of detected syllables could be also a lack of memorization.

2. *The exploration of the harmonic space.* The exploratory phase is apparently free of any biasing element, in the sense that no hint was given to the subjects about the task to accomplish. It seems that, in such conditions, the shape on the responsive floor (the circular ring) has been the most important cue to follow for the subject. Indeed, in nearly all the recorded paths at least one circular path was found, whereas not in all the recorded paths

| **Learning Test General Results** ($1^{st}$ **session**) | | | |
|---|---|---|---|
| Test/Interview/Likert-scale | Not Musicians | Musicians | Difference |
| Syllables of Change | 3.6 (6.4) | 5.5 (4.5) | + 21% |
| Harmonization (CH + PH) | 1 (4.5%) | 5 (22.7%) | + 18% |
| Chords changes | 4.8 (5.2) | 7 (3) | + 22% |
| Chords employed | 3.5 (0.5) | 4.2 (1.2) | - 24% |
| Harmonizing route<br>(more than three chords detected) | 0 | 54% | + 54% |
| Pleasantness | 3.4 (68%) | 3.4 (68%) | 0 |
| Utility | 2.9 (58%) | 3.6 (72%) | + 14% |

**Table 7.10:** *Table with the general results of the* $1^{st}$ *session of Harmonic Walk learning test. Separate data for Musicians and Not Musicians with their performance differences are reported.*

a recursive path was found. Moreover, many of the subjects could not fully understand the difference between the straight path and the circular one, and, consequently, showed a clear tendency in using them in the same way, that's is following the shape's contour. On the other hand, the straight logic of this behavior must be recognized and seriously considered in the future work of refinement of the application design. As already outlined, when following passively the ring, many subjects missed the possibility of building a perceptual and a sensorimotor map of the strongest tonality harmonic progressions, thus compromising the possibility of success in the third phase of the test. The search of a satisfactory harmonization path has nothing to do with a random search, but follows very clear, perceptually strong, recursive tracks. Nevertheless, one of the aims of Harmonic Walk is to encourage the users to discover such features through a perceptual and movement practice.

3. *The partial or full harmonization task* was accomplished by the 27.2% of the twenty-two subjects. Considering the high level of complexity of the melody harmonization this result is not so bad, mainly when no explicit information about melody harmonization has been previously delivered to the subjects.

4. *The number of chord changes* is directly related to the detection of the syllables of harmonic changes, as the number of changes is exactly corresponding to the number of syllables detected. Basically the question is the same, but is put in a different perspective: It seems that the subjects did not realize the similarity of the two questions, as the answers from the same subject differ in all the cases but one. Only subject 7 among the Musicians group gave both correct answers, while subject 9 gave the most coherent answer having detected only four syllables of change but also four chord changes. Furthermore, comparing the records of the successful or partially successful subjects in the melody harmonization task (starred and double starred subjects) it possible to observe that the correlation is scarce or null. Only subject 2 of the Musicians group achieved a good record in the number of chord changes, while all the others made a more or less mistakes. Similarly, comparing the number of different employed chords (three in total, I, V and vi degree) it is possible

to see that there is a good match only with subjects 5 and 6 of the Musicians group, while the other four starred subjects declared all the wrong number of chords. The results of the syllables column are particularly touching and need further investigations, mainly for the complete successful subjects. How can a successful harmonizer miss to detect the syllables of the song for which s/he actually searched and successfully found the right chord? It seems that musical consciousness lays far away from the embodied knowledge employed by these successful subjects. On the other hand, subject 7 of the group of Musicians detects perfectly all the syllables of change (ten) and the number of harmonic changes (ten), but misses to find the number of employed chords (five instead of three); nevertheless s/he fails to harmonize the melody in the given time.

5. *Qualitative assessment results* show that the psychological and cognitive weight of the test's experience has been remarkable, as witnessed by the rich and complex output obtained from the participants. It seems that, independently of the explicit knowledge they succeed to realize at the end of the experience, the Harmonic Walk is able to offer to the users a deeper insight of what a musical experience is. The application's use is not a fatigue, as the average of ratings about its pleasantness is 3.4. In general, the main perceived features upon which the application is built are musical memory, rhythmic skills and sensorimotor reactions that are important factor in musical activities. As a matter of facts Harmonic Walk, in spite of its simple interaction modality, conveys in a very short time a great deal of information about complex music composition processes and this can embarrass the user who is not properly prepared. On the other hand, this is just the field where a didactic approach and careful training can help. Moreover, a most insightful address to the Harmonic Walk can also highlight the great potential of its creative use, which is implicit in the design.

6. *Musicians versus Not Musicians results.* In general, in the quantitative assessment's records, Musicians obtained better results than Non-Musicians, as can be seen at glance in Table 7.10. The number of successful harmonizers was one among Musicians as well as among Not Musicians, whereas no partially successful harmonizer was found among Not Musicians. Looking at Table 7.6 it is possible to see that Musicians were much better than Not Musicians in identifying the right song chord sequence, as the majority of them (54%) could identify a right sequence of more than three chords, while all the sample of Not Musicians equals or remains above this number. In general, in Likert-scale surveys Musicians quoted better the application's utility whereas Not Musicians quoted better its pleasantness. In particular, in the application perception opinion question, two Not Musicians judged it useless, showing that the mere test's experience cannot be considered self-explanatory for all the subjects.

## 7.2 Learning in the Harmonic Walk Environment (Second Session)

Also the second session of the test was organized at the Catholic Institute Barbarigo in Padova (Italy) in January 2015. The second session of the Harmonic Walk learning test aimed at measuring the impact of explicit information on the students performance with respect to four different subject categories selected among high school Musicians and Not Musicians students (see Table 7.11). One half of the participants could get some information about the application's content in a hour demonstrative lesson, while the other half not. The results comparison is presented and discussed with respect to the success in the main application's task (the melody harmonization) and to the level of knowledge acquired in the three selected aspects of the application's musical content (the detection of the syllables of change, the number of harmonic changes and the number of employed chords).

### 7.2.1 Subjects

The subjects of the second session of the test are the same as the first. In the second test the group of Musicians and Not Musicians was randomly divided into two subgroups of five and six subjects. Then, one subgroup of Musicians and one subgroup of Not Musicians was selected to assist to one hour demonstration lesson. The two subgroups are named respectively "instructed Musicians" and "instructed Not Musicians". The four groups of subjects and denominations are defined as shown in Table 7.11.

| | |
|---|---|
| instructed Non Musicians (iNM) | 5 subjects |
| instructed Musicians (iM) | 6 subjects |
| uninstructed Non Musicians (uNM) | 6 subjects |
| uninstructed Musicians (uM) | 5 subjects |

**Table 7.11:** *Subject distribution among the four groups for the second session Harmonic Walk learning test.*

### 7.2.2 Materials, Scenario and Method

The materials and the scenario are exactly the same as in the first session test. The test is organized in the same three phases as explained Section 7.1.4. Only the method differs for instructed and uninstructed subjects in the following way:

1. the eleven chosen subjects take part to one hour demonstrative lesson where the test conductor and the teacher explain the aim of the test and show how the applications works. The conductor shows the meaning of the three phases of the test and how the required tasks could be accomplished (only for instructed subjects)

2. each subject undergoes the three phases of the test, lasting respectively a maximum of five, three and five minutes

3. after the time is expired or the last task is accomplished, the student fulfills a new questionnaire identical to that of the first test.

Instructed subjects follow the test schedule from the beginning, while uninstructed subjects skip point 1.

### 7.2.3 Meta-Analysis Methods

The first session test results are considered as the control and the second session test results as the treatment.[3] After collecting the two sessions test's records about the complete (CH), partial (PH) and unsuccessful harmonization (NH), a 2x3 contingency table is obtained for each of the 4 subject categories (see Table 7.12). For each table the Fisher's exact probability (P-value)[4] is calculated, in order to express the degree of statistically significant association between the control and treatment results, assuming a P-value $< 0.05$ as the significance threshold. Effect size *r*[5] and Cohen's *d*[6] for the data are calculated on the basis of mean and standard deviation for the syllables of harmonic changes, number of harmonic changes and of employed chords in the harmonization task (see respectively Tables 7.13, 7.14 and 7.15).

### 7.2.4 Data Meta-Analysis

In this Section the collected data are analyzed and organized in four items and tables, depending on the musical content. For each assessed ability, data interpretation based on the appropriate meta-analysis methods is provided. The second session assessment test concerned the harmonization task, the syllables of change detection, the number of harmonic changes and the number of employed chords.

1. *The harmonization task* contingency tables (see Table 7.12), show statistically significant improvement results only in the category of instructed subjects (Musicians and Not Musicians), with a very good result for the category of instructed Musicians (iM's P-value = 0.007), while uninstructed Musicians and Not Musicians are well beyond the significance threshold (respectively, P-value = 1 and 0.437).

[3] Usually, in learning environment analysis, the control is given by the results obtained during a traditional lesson, without the help of any technology, while the treatment is given by the results obtained during sessions where the application is used to convey the same contents. In the case of the *Harmonic Walk* this was an impractical assessment condition. Indeed the melody harmonization is a rather difficult task, which requires a great amount of practice and musical knowledge, well beyond the level of a music high school student.

[4] The Fisher's test is used in the analysis of contingency tables when sample sizes are small. Its results are always exact also if the frequency of values is less than five (the frequency validity limit for Chi-test).

[5] Effect-size *r* correlation measures the magnitude of a treatment effect.

[6] Cohen's d is the difference between two means divided by the pooled standard deviation. d=0.2 is to be considered a small effect size, 0.5 represents a medium effect size and 0.8 a large effect size.

**Harmonization task**

| | $1^{st}$ SESSION | | | $2^{nd}$ SESSION | | | P-value |
|---|---|---|---|---|---|---|---|
| | CH | PH | NH | CH | PH | NH | |
| **iNM** | 0 | 0 | 5 | 4 | 0 | 1 | 0.047 |
| **iM** | 1 | 2 | 3 | 6 | 0 | 0 | 0.007 |
| **uNM** | 1 | 0 | 5 | 0 | 1 | 5 | *1.000* |
| **uM** | 0 | 2 | 3 | 2 | 2 | 1 | *0.437* |

**Table 7.12:** *Contingency tables of first and session test results for the harmonization task for each of the four subject groups. CH is the number of subjects who completed the harmonization, PH is the number of subjects who partially completed the task and NH is the number of non successful subjects.*

2. *Syllables of change detection.* Table 7.13 shows rather uniform results among the various subject categories for the detection of the syllables of change, as all Cohen's *d*s are comprised in the same range of values, with a small improvement for all the four subject categories.

**Syllables of change**

| | $1^{st}$ SESSION | | $2^{nd}$ SESSION | | Effect size *r* | Cohen's *d* |
|---|---|---|---|---|---|---|
| | Mean | SD | Mean | SD | | |
| **iNM** | 4.25 | 2.872 | 5.2 | 3.114 | 0.156 | 0.317 |
| **iM** | 5 | 2 | 6 | 3.464 | 0.174 | 0.353 |
| **uNM** | 3 | 2.944 | 3.75 | 1.258 | 0.163 | 0.331 |
| **uM** | 6 | 3.082 | 7 | 1.825 | 0.193 | 0.394 |

**Table 7.13:** *Table of mean, standard deviation, effect size and Cohen's d of the results of the first and second session test for the number of correct syllables of change detected by the four subject categories.*

3. *Number of chord changes.* The results for the detection of the right number of chord changes show a large improved level for the category of instructed subjects, while for uninstructed subjects there is a small improvement ($d$ = 0.176), if not a negative effect in the category of uninstructed Non Musicians.

4. *Number of employed chords.* The results for the number of employed chords show a good improvement in the categories of Non Musicians (both instructed and uninstructed), whereas a negative value for instructed Musicians and a very small effect size for uninstructed Musicians is recorded ($d$ = 0.054).

**Number of Chord Changes**

| | $1^{st}$ SESSION | | $2^{nd}$ SESSION | | Effect size *r* | Cohen's *d* |
|---|---|---|---|---|---|---|
| | Mean | SD | Mean | SD | | |
| **iNM** | 5.5 | 2.121 | 9.2 | 1.095 | 0.738 | 2.192 |
| **iM** | 7 | 2.549 | 10 | 0.408 | 0.634 | 1.643 |
| **uNM** | 4.3 | 1.528 | 4 | 0.816 | *-0.121* | *-0.244* |
| **uM** | 8.25 | 2.872 | 8.8 | 3.347 | 0.087 | 0.176 |

**Table 7.14:** *Table of mean, standard deviation, effect size and Cohen's d of the results of the first and second session test for the number of correct chord changes detected by the four subject categories.*

**Number of Employed Chords**

| | $1^{st}$ TEST | | $2^{nd}$ TEST | | Effect size *r* | Cohen's *d* |
|---|---|---|---|---|---|---|
| | Mean | SD | Mean | SD | | |
| **iNM** | 3.67 | 1.527 | 3 | 0 | 0.296 | 0.620 |
| **iM** | 3.4 | 1.673 | 3.75 | 0.957 | *-0.127* | *-0.258* |
| **uNM** | 5.4 | 3.714 | 3 | 0 | 0.415 | 0.913 |
| **uM** | 3.3 | 1.204 | 3.25 | 0.5 | 0.027 | 0.054 |

**Table 7.15:** *Table of mean, standard deviation, effect size and Cohen's d of the results of the first and second session test for the number of employed chords detected by the four subject categories.*

### 7.2.5 Second Session Test Results Discussion

The twenty-two test subjects were submitted to various assessment conditions to discover what is the weight of the simple test repetition (for uninstructed Musicians and Not Musicians) and of the test repetition after one hour lesson, where the test's musical content was explained and showed by the test conductor (instructed Musicians and Not Musicians).

1. *Harmonization task.* The test results show that, for the harmonization task, the lesson was very important, as instructed subjects (Musicians and Not Musicians) could achieve a very good result in the second session test's session, while uninstructed subjects performed rather poorly. Anyway, the lesson did not provide any technical detail about the harmonization task, but, rather, showed to the subjects how it could be practically done in the application's environment. Thus, the necessary information was transmitted to the subjects only by the observation of the test conductor's movements and interaction style and not by theoretical explanations, thereby proving the high power of knowledge communication of the Harmonic Walk learning environment.

2. *Musical features awareness.* The results of the harmonization task do not always coincide

with the findings in the three musical content knowledge acquisition tests. For instance, in the detection of the syllables of change a general uniform improvement among the four subject categories, ranging from a minimum of Cohen's *d* value of 0.317 for instructed Not Musicians, to a maximum of 0.394 for uninstructed Musicians is observed. It is a small effect size, but it indicates a clear improvement in the musical knowledge, independently from the success in the harmonization task. The explanation of such a result can simply be the memorization of the song's words, which improved with a second trial. Anyhow, the information was explicitly provided to the instructed group during the lesson, but this didn't seem to affect significantly the test's results. The number of harmonic changes is again much better detected by the instructed group, with great Cohen's *d* values. This information, explicitly provided during the lesson, was clearly much easier to remember than the list of the syllables of the harmonic changes and, consequently, most of the instructed subjects could produce a better answer in the second session test. In the perception of the number of the tonal functions employed in the song's first phrase harmonization, a right answer in all the Not Musicians subjects is recorded. This is probably due to suggestions shared among the group components, and, consequently, this part of the test cannot be considered valid.

## 7.3 Movements in the Harmonic Walk Environment

Harmonic Walk exploits the connection between body movements and musical concepts to allow users to practice high-level structured elements like tonal harmony in a simple and effective way. As Harmonic Walk is a music learning application, full-body motion can be regarded not only as the leading factor for human-computer interaction in motion-based environments, but also as a learning modality. The aim of this Section is to offer some elements to understand the relationship between motion and the cognitive tasks required by Harmonic Walk. The observations reported were collected in different experimental sessions conducted in elementary schools (2014), high schools (2014, 2015) and public demonstrations (2014).

### 7.3.1 The three User's Tasks

The Harmonic Walk's ultimate aim is to drive the user towards a tonal melody harmonization, which is a complex multi-faceted task for the user. The goal of melody harmonization can engage the user in the phases and tasks already described in Figure 7.1, that are the harmonic rhythm detection, the spatial memory of chord position and the chord route selection and timing. For each of these tasks the interaction modality is analyzed and discussed.

#### 7.3.1.1 Harmonic rhythm interaction.

The harmonic rhythm interaction requires the user to make a step forward when s/he detects a change of the harmonic region. In general, the resulting observed movement quality is stiffness, which is that users relate to this musical feature in a rather unnatural way. This depends much on the musical quality of the composition employed for the test. Users who detected the harmonic rhythm of the first four bars of Beethoven's Fifth symphony, where two harmonic changes are both marked by a musical fermata (see Figure 5.5), were much easier in the step. As a matter of facts, in these two points the synchronization required is much easier because the fermatas allow a longer release time with respect to common synchronization times.

#### 7.3.1.2 Harmonic space exploration.

During the exploratory phase of Harmonic Walk tests, the following behaviors were observed:

1. Running along the circular ring. Five- or six-year-old children were attracted to the sound produced by the circular ring and simply ran continuously either clockwise and counter-clockwise to hear the effect. This observation shows the great communicative power of interactive environments, which require deep analysis and clearly targeted design strategies in very young subjects.

2. Walking along the circular ring. Elder and high school subjects, when asked to explore the environment, simply walked along the ring. This behavior expresses the high biasing power of the chord arrangement shape. Harmonic progressions have hierarchical relationships, which are not represented in the chord arrangement. Thus, the user may assume that following the circle would lead to some useful musical information about the harmonic space, but soon realizes that it is not so.

3. Performing some elementary harmonic progression. Some user, frustrated by the unsatisfying musical result of the simple circular walk, tried other exploratory strategies by linking only two or three chords at a time. This is a much more fruitful approach because involves an important exploratory element: listening. While listening, the user can apply her/his harmonic implicit knowledge that can drive her/his movements. Listening has a great influence on the movement speed and direction during the exploratory phase, because the user has to synchronize her/his movements to what s/he perceives from the external environment. Thus, if some strong harmonic relationship is perceived, this makes her/him stop there and repeat the route.

4. Exploring newfound progressions following a regular rhythm. The harmonic progression route repetition makes the user always more and more sure about the route to follow, and this allows some metrical entrainment to emerge. Harmonic progressions organized into metrical structure are the premise of a new creative behavior, because they allow new, richer, and original chord progressions to be performed. These can be the basis of new tonal music compositions completely originated from bodily movements.

#### 7.3.1.3 Melody harmonization.

The melody harmonization task requires to the user to perform a true choreography, with a fluent and sometimes elegant but highly constrained motion among the various chord locations. Every user exhibits her/his own motion style while performing the song harmonization, as observed by Holland et al. (2009) in similar conditions. Indeed, it seems that, in spite of its motion constraints, the application is able to evoke in the users the long-term sensorimotor information that characterizes personal movement qualities. Further observations during the tests show that the ability of completing the melody harmonization task builds upon musical memory (i.e., the subject's ability to remember the song, usually expressed by her/his ability to sing it) and rhythmic, regular movements during the exploratory phase. Musical memory is an important guide because it provides users with handy, basic information to direct movement. Moreover, users who had a controlled exploratory phase with slow, rhythmic movements were more likely to complete the melody harmonization task.

### 7.3.2 Movement Analysis

1. *Harmonic rhythm interaction.* Harmonic Walk is based on a physical floor interface. As the name suggests, the user interacts with the application's content through her/his steps, which must be regarded as the only human-computer interaction modality in this environment. Kinematic patterns of adult male human gait (Pietraszewski et al., 2012) report a stride cycle time (two steps) ranging from a minimum of 0.93 seconds for high speed gait to a maximum of 1.26 seconds for low speed gait, with a preferred stride time about 1 second, corresponding to a musical speed of 60 BPM (Beats per Minute) for two steps and to a speed of 120 BPM for the single step. Thus, these measurements can be considered as the optimal step-beat speed interaction. The step onward synchronized with the harmonic rhythm is triggered by the feeling of the upcoming end of the present harmonic region. The user is not worried about where to move, as s/he has in front a row of tagged positions, but rather about when to move. One of the difficulties of synchronizing the human step with the harmonic rhythm may be that usually this is much slower than the average stride speed. Imagining the period of Figure 5.11 as a part of a song played at 100 BPM in a binary musical meter (two beats per bar), the harmonic region durations is 1.2 seconds and 2.4 seconds, longer than the preferred average stride and, for the longest durations, even longer the lowest stride limit (1.26 seconds). However, the embodied knowledge deriving from such step-harmonic rhythm coordination is a very convincing experience of what harmonic rhythm is. For instance, the guitarist left hand's movements on the guitar fretboard has exactly the same entrainment rhythm, because from this movement depend the chord changes. It is a common experience how, when singing together with a guitar accompaniment, the points of harmonic change are marked by head and body movements and how they correspond to marked accents in the right hand's strings percussion. This is a very strong feeling, which improves group performance participation and enjoyment. It is also important to notice that playing the guitar implies the coordination between the left hand (harmonic rhythm) and the right hand (beat-based rhythmic pattern), and that this entrainment action links a lower level of rhythmic pattern organization to the higher rhythmic level of harmonic structure. This two-level rhythmic performance could help also in the case, very frequent in popular music, when the change of the harmonic rhythm does not correspond to the phrases and semi-phrases organization, as shown in Figure 5.11, where

the star points out the overlapping of the two musical structures. The phrase change invites the user to mark it with a step, while the harmonic rhythm would require no change at all. Nevertheless, it is important to prepare the Harmonic Walk users to perform the harmonic changes through step interaction and to train them to move on the floor to accomplish the melody harmonization task. Employing a simple gesture to mark the harmonic changes would not take in account that the harmonic changes have to happen occupying different interactive landmarks with precise spatial relationships, which must be reached through steps inside the circular chord mask.

2. *Harmonic space exploration.* The knowledge of the harmonic space is a fundamental prerequisite for melody harmonization. What is required to the user of Harmonic Walk is not a theoretical knowledge, but an embodied, tacit knowledge that can link her/his previously acquired harmonic experience with the chords s/he hears when occupying one of the six zones of the application's circular ring mask. Thus, the user is assumed to employ her/his perceptual skills to build a cognitive map of the tonality harmonic space by freely exploring the six chord zones. The freedom in this exploratory phase refers to the lack of external rhythmic constraint deriving from the system's audio feedback. The interesting fact in the exploratory movement analysis is that motion schemas are driven only by ideas, emotions and need of confirmation the user feels while moving in the environment. The course s/he draws during the exploration is like a journal of how her/his brain is working to build the harmonic space cognitive map, and, consequently, many exploratory strategies may be found. One of these is the experience of the harmonic turnaround, which is an interesting model of recursive harmonic progression. The turn can be more or less long: however the turnaround returns always on the tonic chord and this repetition, together with an iterative harmonic rhythm, makes it a kind of ever lasting ground used to build melodic variations and refrains. Some turnarounds trajectories are depicted in Figure 7.5, which shows how they are all concentrated around the V and I degree final cadence. This concentration is the expression of the dominant (V degree) and tonic (I degree) hierarchical role in the tonal harmony. Turnarounds are good movement patterns because repetition reinforces harmonic functions perception and chords spatial relationships. Moreover, turnarounds require that route segments among the various chords positions are covered with regularity. As harmonic turnarounds are no more than extended patterns of harmonic progressions, the embodied knowledge deriving from this practice is a very useful background for melody harmonization.

3. *Melody harmonization.* Actually, to perform a good song harmonization the user of Harmonic Walk has not only to have a clear idea of when to move, but also of where to go, exactly like a dancer who has to follow a previously appointed choreography. The information of where to go is included in the cognitive map the user has to build during the environment exploratory phase, where s/he has to link the chord sounds with the chord location. Nevertheless, to succeed in the task implies also other important abilities like musical, spatial, and route memory, physical awareness of beat, body movements control, self-confidence, gait fluency, and capacity of getting into the game. Moreover, the

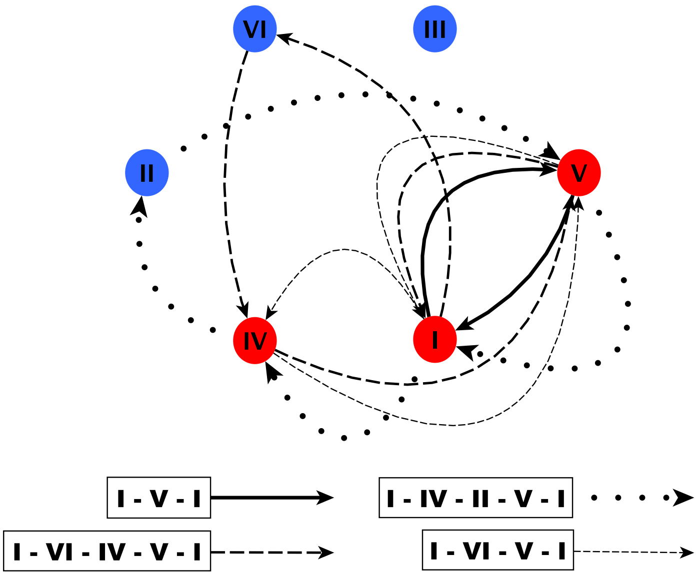



**Figure 7.5:** *Spatial representation of four harmonic turnarounds on the Harmonic Walk interface, with the main chords in red and the parallel in blue. The turnarounds trajectories show the concentration of movement around the V and I degree.*

best results were observed when melody harmonization is the outcome of a collaborative work, i.e., when a group sings the song and one from the group in turn tries to harmonize it jumping from one position to the other. Cooperation offers the advantage of sharing the user's cognitive load of remembering and singing the song with the group, and, in the same time, engages the user in the time constraints with the group's singing. This helps the user's effort in achieving the task and rewards her/him in case of melody harmonization completion.

## 7.4 Learning Test Conclusion

At the beginning of this Chapter some questions were posed about the learning power of bodily interaction. Outstanding elements emerging from the two session test's results are now discussed with respect to them.

### 7.4.1 How does Bodily Interaction Relate to Abstract Knowledge?

As explained in Section 5.1.3, Harmonic Walk's abstract knowledge is expressed through the application's interface design, which is shaped following the spatial representation of the harmonic regions sequence (Section 5.1.4) and of the major tonality harmonic space (Section 5.1.5). However, the two representations require a different level of bodily interaction, one involving movement coordination to the harmonic rhythm and the other involving also the use of the musical chords cognitive map. These two elements form the paradigm of interaction described in Section 4.3 and represent the means through which harmonic abstract knowledge becomes a bodily experience. The test's results, as well as direct observation, show that the circular ring shape has a strong biasing power on users movements. The problem of the harmonic space interface lies in the fact that the musical chords have not all the same importance, as the interface seems to suggest. Interactive graphics, as well as perceptual training based on the practice of the most common turnarounds, may help users in developing a sensorimotor map useful to understand tonal music chords hierarchy and functions.

### 7.4.2 Is it possible to make Embodied Knowledge become Explicit?

The results obtained in the first session test show that a little explicit knowledge emerges after a first experience in the Harmonic Walk environment. After the second session test a small level of improvement for the detection of the syllables of change for all the four categories and a good level of improvement for the number of chord changes in the instructed Musicians and Not Musicians categories is recorded. Thus, in general the results do not show a clear influence of the treatment on the subjects performance. However, a certain increase of musical knowledge emerges after the experience repetition. The most outstanding finding is that the success in the harmonization task does not coincide with the musical elements awareness. As pinpointed in Section 7.1.6.5, the subjects who completed the harmonization task in the first session test missed the correct reconstruction of the harmonizing route. Similarly, the great improvement in the second session test of the instructed group in the harmonization task (see Table 7.12), is not accompanied by the same level of improvement in the syllables of change detection (see Table 7.13).

### 7.4.3 Is Enactive Learning more effective than Explicit Information?

Movement imitation has been crucial in the performances registered for the harmonization task in the second session test. The mere observation of the movements of the test conductor while performing the harmonization task has been able to transmit only by imitation and in a very short time a lot of information to the test subjects. As a matter of facts, the only clear effect of the treatment is the improvement of the instructed group in the harmonization task, as shown in Table 7.12, where it is clear that the lack of a practical example of melody harmonization makes the difference. Many other information has been delivered during the one hour lesson administered to the instructed Musicians and Not Musicians group. However, explicit information has proved not as effective as movement observation and enactive learning.

# Chapter 8

# Harmonic Walk Interface Test

Two different tests are presented in this Chapter under the common label of Harmonic Walk interface test. The first test concerns the effects on user's performance of the chord spatial arrangement on the Harmonic Walk active floor surface. Two arrangements are proposed and the user's performance is measured under the two different conditions. The second test compares user's performance and preferences with respect to the song harmonization task on the active floor surface and on a touch screen. The results show that the physical effort implied in full-body interaction plays an important role in movement pattern memorization and that, in spite of the playful approach proposed by the active floor interaction modality, users perform better on the touch screen and prefer this latter to full-body interaction. These results are very important for the design of interactive spaces applications, as they offer meaningful perspectives regarding the role of full-body interaction in these environments.

## 8.1 Harmonic Walk Chord Arrangement Test

The Harmonic Walk chord arrangement test aims at verifying the effects of chord spatial arrangement on user's performance. Two arrangements are tested. Arrangement 1 derives from the tonnetz's tonal harmony spatial representation discussed in Section 5.1.5, which is an abstract spatial rendering of the chord relationships grounded on renowned theoretical and perceptual basis. Arrangement 2 is built following opposite criteria than those arising from the tonnetz. In arrangement 1 the three main roots (in red in Figure 8.1) are grouped in a half of the circular ring, and the parallel roots (in blue in Figure 8.1) in the other half. The main and parallel harmonies are well separated and each parallel root is put in front of its main as shown by the white lines. Arrangement 2 on the contrary disrupts completely the circular ring division between main and parallel roots and the position's reciprocity of a main root with its parallel. Moreover the most used chord progression (I-V degree) has the shortest step length in arrangement 1 and a medium step length in arrangement 2. The experimental hypothesis is that the group of subjects trained with arrangement 1 should benefit of its spatial and perceptual qualities and thus perform better that the group trained with arrangement 2.

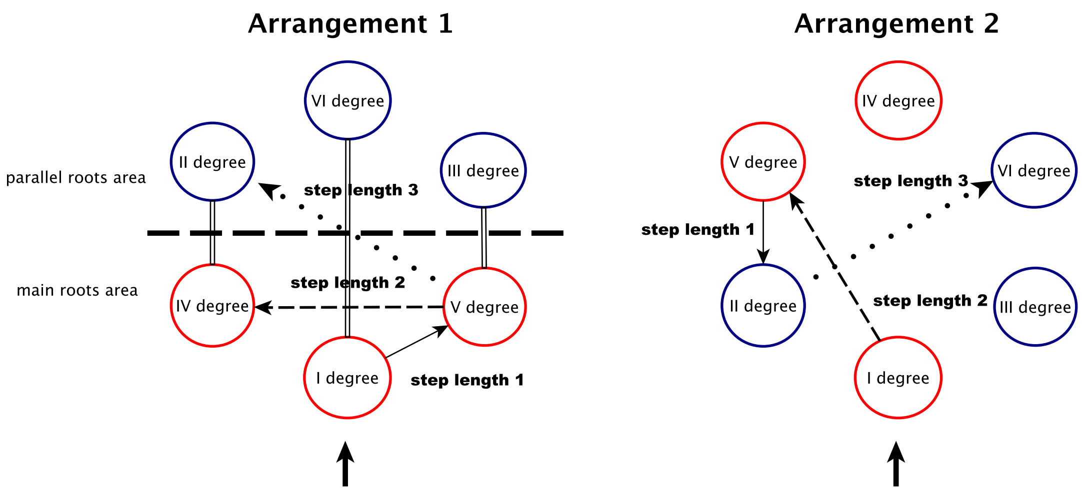



**Figure 8.1:** *The two chord arrangements employed for the Harmonic Walk interface test. The red circles are the main roots and the blue are the parallel. The straight, dashed and spotted arrows indicate respectively the shortest step lenght (1), the medium (2) and the longest (3).*

### 8.1.1 Subjects

A total number of thirty-two elementary school girls and boys between ten and eleven years old took part in the test. The students are equally subdivided into two different groups of sixteen subjects each: one group is trained with arrangement 1 and the other with arrangement 2. All subjects have no musical instrument training in their school curriculum.

### 8.1.2 Materials

The test is based on a simple children song. To guarantee a uniform level of song knowledge from the subjects, the song has been expressly composed for the test. The required musical characteristics of the song are the following:

a) the melodic movements must be reduced to the minimum;

b) the harmonic rhythm must be regular, without harmonic syncopations;

c) the harmonic progression must include at least one main chord (beyond the first degree) and one parallel chord;

d) the melody should employ preferably repeated patterns;

e) the melodies must be written in a major tonality, without modulations or extraneous pitches.

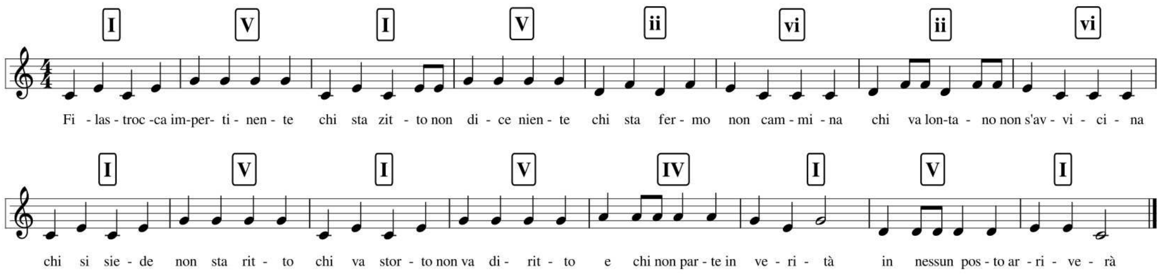



**Figure 8.2:** *The melody employed for the Harmonic Walk chord arrangement test. The song text is reported together with the harmonic changes.*

| Song semi-phrases | SP 1 | | | | SP 2 | | | | SP 3 | | | | SP 4 | | | | TOTAL EFFORT |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Harmonic Changes | I | V | I | V | ii | vi | ii | vi | I | V | I | V | IV | I | V | I | |
| Step length arr. 1 | - | 1 | 1 | 1 | 3 | 1 | 1 | 1 | 3 | 1 | 1 | 1 | 2 | 1 | 1 | 1 | 20 |
| Step length arr. 2 | - | 2 | 2 | 2 | 1 | 3 | 3 | 3 | 2 | 2 | 2 | 2 | 1 | 3 | 2 | 2 | 32 |

**Table 8.1:** *Table of user's efforts in harmonizing the melody for the Harmonic Walk chord arrangement test.*

Three proposed melodies have been analyzed by four musicians experienced in test management. The melody chosen for the test is depicted at Figure 8.2 where the song's text and harmonic changes are shown. The song's harmonic rhythm requires one harmonic change for each of the sixteen song bars. The changes are regular and are organized in group of four bars (musical semi-phrases, SP from now on). SP 1 is I-V-I-V with all the harmonies in the main root area. SP 2 is ii-vi-ii-vi with harmonies in the parallel root area, SP 3 is again like SP 1 and SP 4 is IV-I-V-I with harmonies in the main root area. To obtain a measure of the overall user's physical effort spent in arrangement 1 and 2, a number related to the step lengths involved in the two arrangements is provided at Table 8.1. With reference to Figure 8.1, step lengths are calculated as 1 (straight arrow, shortest length), 2 (dashed arrow, medium length) and 3 (spotted arrow, longest length). The step length calculation results report an overall effort of 20 for arrangement 1 and of 32 for arrangement 2, with a difference of about the 30% for the benefit of arrangement 1. Moreover, the step lengths occurrences shown in arrangement 1 are disrupted in arrangement 2. Particularly, the change of harmonic zone is marked by a 3 or 2 in arrangement 1. The zone change is followed by a series of 1, which points out that when a zone is reached the movements inside it are as short as possible.

### 8.1.3 Method

The control group is assigned arrangement 1 and the treatment group is assigned arrangement 2. Preliminary tests have been organized to verify if the chosen melody fits the tests requirements, and if the test procedure has the right event sequence and timing. Moreover, two experts performances have been recorded to be employed as mean reference times for subjects' performances evaluation. The test is subdivided in three sessions:

1. First session: one hour class lesson about harmony and melody accompaniment. The aim of the lesson is to teach the movement pattern related to the song harmonization to the children. During the lesson the song chosen for the test is used and repeatedly sung by the teacher and pupils. Games on the song's harmonization, chord progressions and harmonic rhythm are proposed. Movement games where students move in time with the harmonic changes are performed following the chord arrangement assigned to each group. The lessons are carried out in separate sessions for group 1 and group 2, because the chord arrangement is different for the two groups.

2. Second session: one hour training in the test laboratory where children learn how to interact with the Harmonic Walk active surface and what is the sound produced by their movements. A touch screen version of the active floor is also presented and tested by the children. The second session is carried out in separate sessions for group 1 and group 2, as for the first session.

3. Final session: individual ten minutes test. The final session of the test is carried out in a laboratory, one subject at a time through a semi-automatized procedure. Only the test conductors are present in the laboratory. The subject is introduced in the test laboratory and is asked to answer a pre-test question about application's pleasantness. A pre-recorded virtual voice drives the subject through the various phases of the test, whose progress is determined by the test conductor. First the subject is invited to explore the environment and listen to the sounds of the various chords. Then the system performs a piano-only version of the song and invites the subject to sing it. At last, when the song has been recalled to the subject's memory, a recorded audio file with only the melody is played while the subject has to harmonize it by moving on the active surface. Three trials are allowed to the subject, which continue until the end independently from the quality of the subject's performance or mistakes. At the end the subject is asked to answer a post-test question about application's pleasantness and to complete a questionnaire (see Section 8.3).

All the sessions of the test have been organized at a time interval of at least fifteen days.

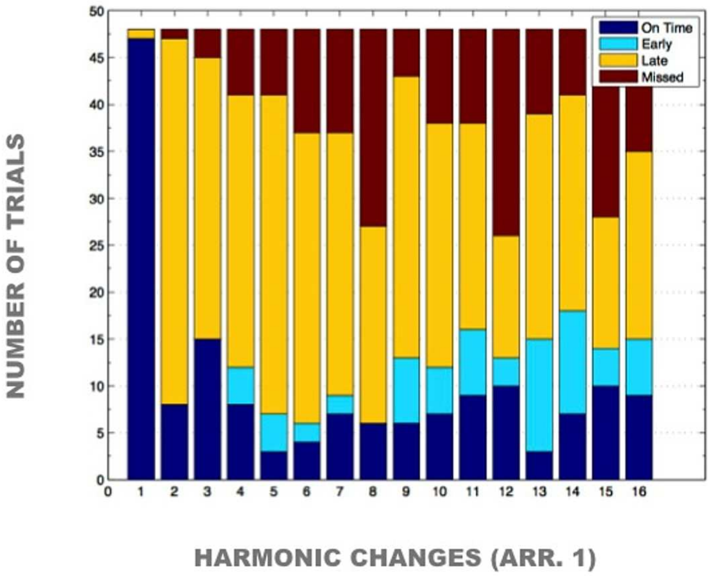
### 8.1.4 Experimental Data

During the test the following data have been collected:

a) the log of the numerical sequence representing the chords played by the subjects;

b) the timing of the harmonic changes.

As every subject performed three trials, there are forty-eight trials for arrangement 1 and forty-seven for arrangement 2 (one has been missed by the system). Figure 8.3 reports a visualization of the number of on time, early, late and missed harmonic changes for each of the sixteen harmonic changes in the two arrangement conditions. These data are employed to determine how



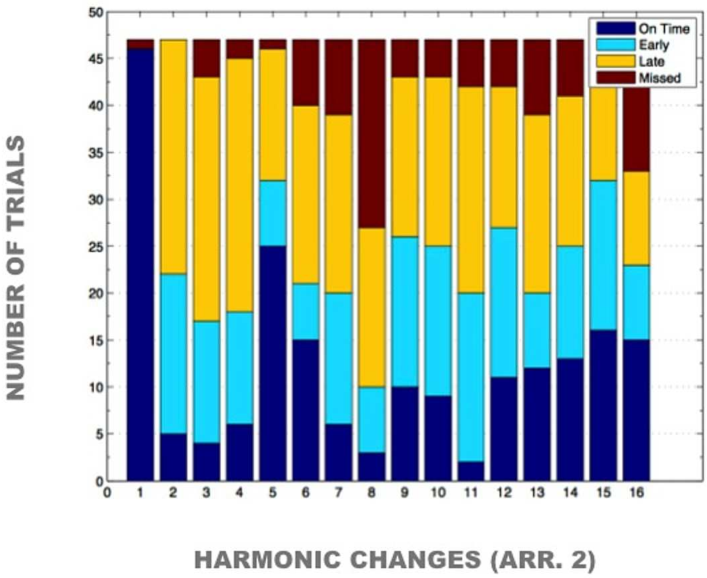




**Figure 8.3:** *The two histograms of the users performances in the two arrangement conditions for each of the sixteen harmonic changes.*

| | | Number of Trials | | |
|---|---|---|---|---|
| Harmonic Changes | | Arrangement 1 | Arrangement 2 | Cohen's $d$ |
| Hit | On time | 142 | 198 | - 0.759 |
| | Late | 400 | 276 | 1.29 |
| | Early | 51 | 186 | - 2.76 |
| **TOTAL HIT** | | 593 | 660 | - 2.57 |
| **TOTAL MISSED** | | 175 | 92 | 0.765 |

**Table 8.2:** *Table reporting the number of trials and Cohen's $d$ of the hit and missed harmonic changes in the two arrangement conditions. For the hits also the subdivision in on time, late and early results is reported.*

much of the trials have recorded an on time, late, early or missed position for each harmonic change. As the steps of the harmonization sequence have not all the same length, for each of the sixteen harmonic changes required for the song harmonization a mean of position timing is calculated on the basis of the reference performances. The minimum and the maximum of each position time is considered as the acceptance range for the on time performance. If the position is reached outside the appointed time range the position is considered or early or late. If the position is not reached at all it is considered missed. The test result are reported in Table 8.2. A Chi-Square test was performed on the independent variables of the total of the hit and missed positions to evaluate the statistical significance of the collected experimental data. The test Chi-Square statistic is 29.21. The $p$-value is 0.00001 and the result is significant at $p < .05$. To evaluate the effect size of the treatment (arrangement 2) the Cohen's $d$ is calculated for each of the harmonic changes categories (on time, late, early and missed and for the total of the hit). As the first chord position does not depend on the subject's choice but was indicated by the system, the related data have been omitted from calculation. The negative Cohen's $d$ values show that the results are contrary to the expected direction.

### 8.1.5 Chord Arrangement Test Results Discussion

Contrary to the experimental hypothesis, the treatment group performed better than the control group with regard both to the total of the hit result (- 2.57) and to the on time hit result (- 0.759). Another great difference is the early hit result (- 2.76). These results suggest that:

1. the perceptual characteristics of arrangement 1 (the division between the main and the parallel root area) had no effect on subjects performances. On the other hand subjects have not been trained to perceive the differences between the two zones but only to remember the harmonization movement pattern;

2. the remarkably less effort required to perform the harmonization task in arrangement 1 did not help the control group to be more successful than the treatment. The only positive control group results are in the late harmonic changes hits (1.29) and in the missed (0.765), which indicate uncertain or unsuccessful performances;

| Song semi-phrases | SP 1 | | | | SP 2 | | | | SP 3 | | | | SP 4 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Harmonic Changes | I | V | I | V | ii | vi | ii | vi | I | V | I | V | IV | I | V | I |
| Step length arr. 2 | - | 2 | 2 | 2 | **1** | 3 | 3 | 3 | 2 | 2 | 2 | 2 | **1** | 3 | 2 | 2 |
| On time Hits (%) | 98 | 11 | 9 | 13 | **53** | 32 | 13 | 6 | 21 | 19 | 4 | 23 | **26** | 28 | 34 | 32 |

**Table 8.3:** *Table of step lengths and of on time hits percentages in harmonizing the melody in arrangement 2 (treatment group). In bold the on time hits percentages in correspondence with the beginning of semi-phrase 2 and 4.*

3. on the contrary, the major physical effort seems to have provoked an increased attention in the treatment group. They recorded a large effect size for early and for on time positions (respectively - 2.76 and - 0.759), which indicate determinate and successful performances;

4. the highest number of on time positions after the first harmonic change is on the first harmonic change of SP 2 (53%), in correspondence of one of the two step length 1. It seems that in this case subjects took advantage of the short step length, whereas this did not happen at SP 4 first harmonic change (in bold in Table 8.3). One possible explanation is that at SP 2 there is a change in the harmonic area (from main to parallel) whereas at SP 4 there is no change.

Thus, the experimental hypothesis is not confirmed by data meta-analysis.

## 8.2 Harmonic Walk Full-body versus Touch Screen Test

This test aims at verifying the effects on subjects performance of two different interaction modalities: the full-body where the user moves on the active floor to harmonize the melody; and the touch where the user performs the same task on a touch screen. For this test only chord arrangement 1 is employed. The interface appears as depicted in Figure 8.4 both in the projection on the active floor and on the touch screen. The experimental hypothesis is that the results obtained trough the interaction in the active floor (from now on called "full-body") are better than the results obtained with the touch screen (from now on called "touch"), as physical effort and full-body interaction are supposed to be more effective in helping the subjects to perform the harmonization movement pattern.

### 8.2.1 Subjects, materials and method

Twelve elementary school girls and boys between ten and eleven years took part in the test. To avoid that the test results are influenced each other, six subjects made first the full-body test and then the touch screen test, whereas the other six subjects made the contrary. Materials and method are the same as in the previous test.

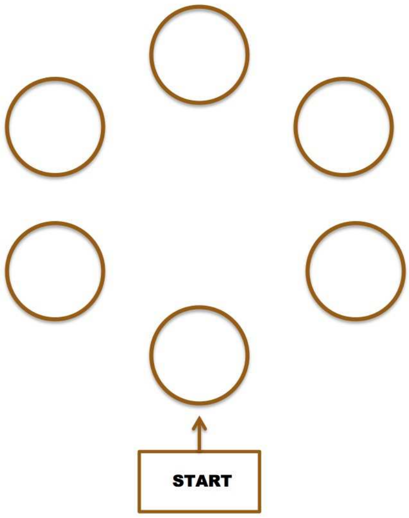


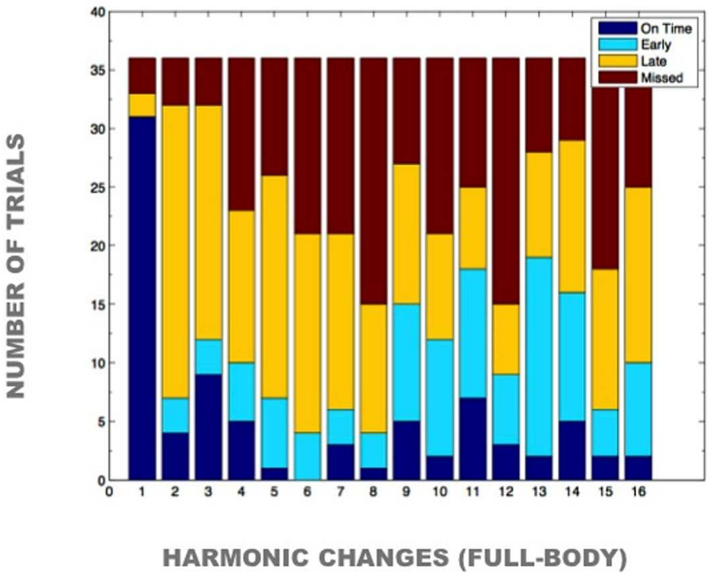

**Figure 8.4:** *The active floor and touch screen interface employed for the Harmonic Walk test.*

### 8.2.2 Experimental Data

As every subject performed three trials, there are thirty-six trials for the full-body and for the touch screen condition. Figure 8.5 reports a visualization of the number of on time, early, late and missed harmonic changes for each of the sixteen harmonic changes in the full-body and in the touch condition. These data are processed in the same way as the previous test to determine the acceptation range for the on time performance. The test result are reported in Table 8.4. A Chi-Square test was performed on the independent variables of the total of the hit and missed positions to evaluate the statistical significance of the collected experimental data. The test Chi-Square statistic is 7.51. The $p$-value is 0.006 and the result is significant at $p < .05$. To evaluate the effect size of the touch screen condition, the Cohen's $d$ is calculated for each of the harmonic changes categories (on time, late, early and missed and for the total of the hit). As the first chord position does not depend on the subject's choice but was indicated by the system, the related data have been omitted from calculation. The negative Cohen's $d$ values show that the results are contrary to the expected direction.

### 8.2.3 Full-body versus Touch Screen Test Results Discussion

Again, contrary to the experimental hypothesis, the touch results are better than the full-body with regard both to the total of the hit result (- 1.15) and to the on time hit result (- 1.01). Again a great difference is reported in the early hits result (- 1.48). These results are very similar to those of the chord arrangement test and suggest that the touch interaction modality produces better results than the full-body in the on time and early harmonic changes, whereas there are less late and missed chord positions than in the full-body. Also in the touch condition there is a negative peak in the missed at SP 3 first harmonic change. In this case there is no step length facility, as the test condition is on the touch screen. Anyhow, it seems that the change from the parallel



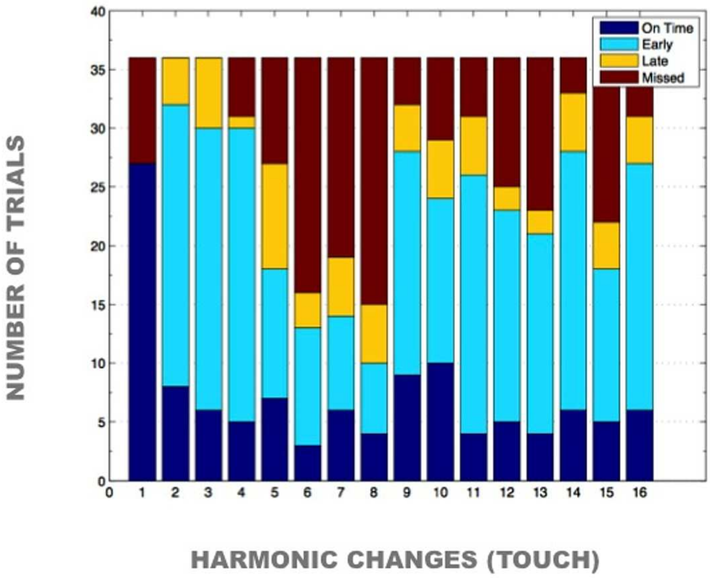




**Figure 8.5:** *The two histograms of the users performances in the full-body and in the touch condition for each of the sixteen harmonic changes.*

| | | Number of Trials | | |
|---|---|---|---|---|
| Harmonic Changes | | Full-body | Touch | Cohen's $d$ |
| Hit | On time | 82 | 115 | - 1.01 |
| | Late | 205 | 64 | 1.02 |
| | Early | 104 | 254 | - 1.48 |
| **TOTAL HIT** | | 391 | 433 | - 1.15 |
| **TOTAL MISSED** | | 185 | 143 | 0.31 |

**Table 8.4:** *Table reporting the number of trials and Cohen's $d$ of the hit and missed harmonic changes in the full-body and touch condition. For the hits also the subdivision in on time, late and early results is reported.*

| Song semi-phrases | SP 1 | | | | SP 2 | | | | SP 3 | | | | SP 4 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Harmonic Changes | I | V | I | V | ii | vi | ii | vi | I | V | I | V | IV | I | V | I |
| Missed Hits Touch (%) | 25 | 0 | 0 | 14 | 25 | 56 | 47 | 58 | **11** | 19 | 13 | 30 | 36 | 8 | 39 | 14 |

**Table 8.5:** *Table of missed hits percentages in harmonizing the melody in the touch interaction modality. In bold the missed hits percentages in correspondence with the beginning of semi-phrase 3.*

harmonic area of SP 2 to the main harmonic area of SP 3 may have produced an increase in the attention and thus a major number of on time hits. Thus, also this experimental hypothesis is not confirmed by data meta-analysis.

## 8.3 Subjects Evaluations

A questionnaire to collect evaluations about the Harmonic Walkwas prepared. The questionnaire includes a pre- and post-test question for all the tested subjects and a question about their willingness to repeat the experience. Only the group that made the full-body and the touch test were asked to answer an additional post-test question about the touch experience and their preference for the full-body or the touch, or of both application's versions. For the pre- and post-test questions a five points emoticons Likert-scale is employed. The pre- and post- test questions concern the subjects evaluation about the application's pleasantness following the scale shown in Figure 8.6. The answers about the subjects' willingness of repeating the experience are evaluated as 1 (no), 2 (perhaps) and 3 (yes). The touch group were also asked if they preferred the touch or the


**Figure 8.6:** *The five point emoticons Likert-scale employed for subjects evaluation in the Harmonic Walk tests.*

| | PRE-TEST FULL-BODY | POST-TEST FULL-BODY | REPETITION | POST-TEST TOUCH | PREFERENCES | | |
|---|---|---|---|---|---|---|---|
| | | | | | Touch | Full-body | Both |
| Control group | 3.7 | 3.9 | 2.6 | | | | |
| Treatment group | 4.2 | 4.2 | 2.9 | | | | |
| Touch | 4.08 | 4.25 | 3 | 4.5 | 50% | 33% | 17% |

**Table 8.6:** *Table with subjects evaluations of the regular, wrong and touch test groups.*

full-body version or both. The answers are reported at Table 8.6. Pre-test question aims at evaluating how the subjects consider the experience in the Harmonic Walk environment relying only on the test's second session, where subjects could explore the environment. The treatment group reports the best mean value (4.2) upon the five point Likert-scale. Post-test question aims at evaluating if the experience has changed this judgement. A week improvement can be observed for the control and touch group, whereas the treatment group record no improvement. The touch group expresses also the best post-test pleasantness evaluation and the best answer for experience repetition. At the end the touch interaction modality obtains the 50% of the preferences among the twelve subjects of the touch group. Thus, in general the application reaches a very good evaluation judgement both before than after the test, as confirmed by the fact that the great majority of the subjects would like to repeat the experience. Nonetheless, the highest degree of pleasantness is for the touch interaction modality, which is preferred also by a half of the touch group subjects.

## 8.4 Harmonic Walk Interface Test Conclusion

The Harmonic Walk Interface test have been organized to verify the impact on users performances of the interface spatial organization and of the interaction modality connected to the active floor or to the touch screen interface. The first experimental hypothesis is that the chord spatial organisation deriving from theoretical and perceptual basis (arrangement 1) works better than an arrangement that completely disrupts arrangement 1 principles and causes a major physical effort to the user (arrangement 2). Experimental data are clearly contrary to the experimental hypothesis as subjects performed much better in arrangement 2 than in arrangement 1, in spite of the major effort required. This result does not refute chord arrangement 1, but rather points out the role of physical effort in subjects memory and performance. It seems that the major effort required has been able to arouse subjects attention on the movement pattern, allowing thus a better result. Nevertheless, also musical perception plays a role, as shown in some cases by the peaks recorded in the changes between the main and the parallel chord area. The second experimental hypothesis concerns the difference between the interaction on the active floor surface and on the touch screen, whereas the first is assumed to be more effective on subjects memory and performance. Also in this case the experimental hypothesis has not been confirmed, because subjects performed a little better in the touch than in the full-body condition. Some subjects affirmed that the touch screen was much easier than the active floor, and that this was the reason of their preference. Undoubtedly the touch screen interaction is more precise than the active floor and, moreover, users in general are much more accustomed to the touch screen rather than to full-body interaction. On the other hand, this results resonate with those obtained by Johnson-Glenberg et al. (2011), who did not get any significant learning improvement in comparing physical interaction in the SMALLab environment and traditional mouse and computer screen interaction.[1]

### 8.4.1 Consequences for Application Design

The experimental results exposed above help to clarify some important design principles for active floors applications. The first test result was completely unexpected and probably represents only a hint of how full-body interaction influences users in environments like Harmonic Walk. The world of full-body human computer interaction claims for discovery, and deeper investigation is still needed to understand how it works. On the other hand touch screen interaction seems to be more successful than full-body interaction, though reactive floors have much more facilities than touch screens. Some of them are the following:

1. *Reactive floors allow a wider audience participation.* Application design has to consider the importance of social interaction as a fundamental design element. The possibility of sharing a common experience with co-located users or with simple by-standers can make an important difference in the learning experience. The common search of problem solutions as well as the sharing of trials and efforts can reinforce cognitive processes and group socialization.

2. *Reactive floors can provide an augmented reality immersive environment.* In the present Harmonic Walk tests the environment responsiveness has been limited to audio and to very basic graphical elements. Nonetheless, the combination of interactive audio and graphics can provide a much powerful immersive experience which can engage users in a much more convincing way.

3. *Reactive floors can become invisible and allow interesting discovery walks.* Contrary to immersiveness, reactive floors can be embedded in public spaces and completely disappear to the user's sight. Thus reactive floors can deploy all the magic of a discovery walk, where users are not aware of being tracked and consequently do not expect any reaction from the system.

Future research on active floor applications needs to be re-defined on the basis of these fundamental emerging points which enhance the characteristics of responsive environments with respect to touch screens.

[1]The SMALLab (Situated Multimedia Art Learning Lab) system has been presented at Section 3.1.2.

# Chapter 9

# Conclusion

The main purpose of the present thesis is to show how abstract knowledge can be conveyed to the user through spatial representations. Interactive spaces, designed according to the landmarks arrangement that derives from musical concept's spatial representation, are employed to allow the user to actually enter the active regions and act inside them. Her/his movements are ruled by precise time and space constraints that derive from the musical structure. Thus, user's movement are the expression of the concepts themselves. They are in the same time the measure of the user's knowledge and the means by which s/he can increase it. In the interactive space everything is mediated by movement and particularly by its direction and timings. The direction depends on the user's spatial cognition and conceptual map of the interactive environment; the timing depends on user's cognition and synchronization to musical events. Five case studies devoted to music learning, active listening and interactive composition and expressive performance are presented. The design principles upon which the music applications are based are summarized in a common general framework, which explains the relationship between the application's musical content and the interactive landmarks or thresholds spatial arrangement. The interaction paradigm focuses on the active surface or three-dimensional regions where the user's moves following the audio and visual feedback produced by the system. Assessment session results show that musical abstract knowledge can actually reach also untrained users and allow them to accomplish complex tasks like melody harmonization. A certain amount of musical knowledge and awareness of the employed musical concepts is achieved by test subjects, also if no additional musical information is provided. Notwithstanding, subjects who could gain some insight about the musical features were not successful in the harmonization task and vice versa, showing that embodied knowledge and musical features awareness seem to travel on different tracks. Moreover, subjects who received both information about the musical features and a demonstration of the movement pattern necessary for the harmonization task improved their performance only with respect to the latter, showing the importance of pattern imitation. Further tests on the Harmonic Walk interface seem to suggest that physical effort also plays a role in movement pattern performance. However, if the same task has to be performed on an active floor and on a touch screen surface, users prefer the latter and perform better in the touch modality than in the full-body. All these data point out that full-body interaction is completely different from other forms of physical interaction like touch or point. It is a powerful interaction means that has its own rules. It is more sensitive to enactive learning than to explicit information; it is influenced by physical effort and it is not necessarily linked to conceptual awareness; it is extremely appealing mainly for young users who consider full-body interaction as one of the most engaging technological games.

## 9.1 Contribution for Music Education

Music is intertwined with movement. The most successful educational practices like the Orff and Dalcroze methods are well aware of this. They employ body and space to convey the learning of important musical abilities like rhythmic performance and musical improvisation. By emphasizing the importance of learning by doing, these methods enhance enactive knowledge as the primary source of music skills. My motion-based music application follow the same principle, but exploit the huge potential of sound digitization and computer vision techniques to reach that part of musical knowledge usually reserved to experts. Music composition and analysis, feature awareness and expressive performance are now available not only for musicians but also for the great public of musically untrained users, offering thus a completely new perspective to music education. My motion-based music applications can be considered in all respects real musical instruments, with the only difference that they do not play single pitches like common acoustic instrument do. They rather play with the representation of interpreted musical concepts like tonal harmony, musical structure or composition algorithms. Their use does not require particular technical abilities because the interaction modalities are basic everyday-life movements. Moreover, they allow not only to discover the musical structure but also to manipulate it in a creative way changing the form and length of a musical piece or modifying a melody harmonization or a harmonic turnaround. Many young music students ignore what music expressive performance is, due to the technical limitations of instrument playing mainly in the first years of their studies. This is a serious drawback to their creativity and musical training, and it is probably one of the causes of the frequent abandonment of musical studies. Expressive music performance means to understand what is important in music communication and to enhance it in the proper way. It is amusing and engaging, and has the power to disclose the performer's personality to others. Hand Composers offers the possibility to train dynamic interactions with musical events in an easy and handy way, whereas Disembodied Voices drives expressive interaction well beyond the classical view of time and dynamic modeling. In fact, it employs digital sound processes whose parameters depend on bodily movements to shape sound in a personal and original way. This is a great opportunity of expression for musicians both at professional and at student level.

## 9.2 Perspectives for Human-Computer Interaction

Current trends of human computer interaction seem to converge on bodily motion as an interesting and unavoidable input modality. Although the most common and functional human-machine interaction modality remains the WIMP paradigm (window, icon, menu and pointer device) in professional activities, it is undeniable that post-WIMP interaction styles have caught on thanks to the massive penetration of mobiles and other hand-held devices. As the debate on the introduction of multi-touch surfaces in personal computers testifies, the immediacy and intuitiveness of tapping, pointing and dragging gestures have now become customary for a large number of users. If this is the current state of art, there are other important factors to consider in order to properly evaluate the role of bodily interaction.

### 9.2.1 Social Sharing of the Digital Experience

The great energy spent every day by millions of people in actions aimed at implementing their digital life is destined to find its social counterweight beyond the sharing of virtual spaces like social networks or distributed workspaces. Especially when looking at social interaction of co-located users, the dimension and position of the active area plays a fundamental role. Thus responsive floors, interactive walls and large touch screen displays have become customary in many public exhibitions and artistic performances together with the connected bodily interaction gestures. In these scenarios the use of the WIMP interaction paradigm acquires an ancillary role or is completely absent.

### 9.2.2 Bodily Interaction and Real Life

Another booster of bodily interaction is the growing interest towards augmented and virtual reality environments. Both approaches meet bodily interaction when trying to reproduce real life situations and strive for immersiveness (virtual reality) or when modifying real life through visual or audio feedback (augmented reality). Bodily interaction is the medium by which people experience the world and thus it is the stronghold of research in these fields. The experience of world has also shaped movement patterns that belong to human sensorimotor culture and that can be employed in virtual or augmented environments. The walk, the tap and the pointing are some basic examples. There are cases where joining hands or rising arms all together are interpreted by the system as a common input shared among a group of co-located users. Moreover, more complex models like that of the musical conductor can be considered when a richer interaction is needed.

### 9.2.3 Bodily Interaction as a Means of Knowledge

Bodily interaction does not only provide input buttons for the interactive system but conveys also a great amount of other available data. In the human walk it is possible to track direction, foot pressure, friction and timing; in body limbs movements direction, angles, velocity and acceleration. The multimodality of bodily interaction input is used in learning environments to express some quality of the application's subject and to play-act its behaviour. Kinematic properties of users' bodily interactions are employed to discover physics elements qualities and to make predictions about them (Enyedy et al., 2015). The user embodies some physics property and learns the element's behaviour through her/his body. The same principle is applied in the Harmonic Walk, where the user has to move in time with the harmonic rhythm and has to reach the right chord position play-acting the same movements that harmonic regions do in the harmonic space.

Other experiments of play-acting applications were made for geometry and trigonometry. A user play-acts the upper vertex of a triangle that corresponds to her/his positions on the responsive floor. The basis of the triangle is still while the user, moving the vertex, can see the effects of changing among the various types of triangles. Or, for trigonometry, a user moves along a sphere diameter acting as a cosine input for interactive sine and angle calculation.

## 9.3 Extended Learning Spaces

In spite of these important facilities and potentials, bodily interaction is still mainly confined to the use of mobile devices and has not yet entered users' everyday life in a meaningful way. One reason of this is the very limited experience in interactive spaces where bodily interaction can be widely tested. Interactive spaces themselves are not so common outside research laboratories, due to the difficulty, costs and technical assistance necessary for their setup. On the other hand, the present thesis test results show how physical movement interacts with cognitive tasks, making the learning experience more effective, but also how physical experience itself seems to travel on an independent track that projects the acquired knowledge towards totally new perspectives. All these elements indicate that it is important to build learning spaces that try to fill the gap between the limited, individual, mouse and keyboard space of personal computers to the large, social, multi-user, full-body interactive space. That is, building for the user the possibility of accessing learning spaces in different situations. For example, after the experience in an interactive social space at school with the assistance of the teacher and with the participation of school mates, a student could refine her/his experience and knowledge with a restricted group of friends in front of a computer screen or on a tablet or alone on her/his smartphone. The individual or group results could be registered on a dedicated web server and displayed for the class in the next working session. This is the direction taken by some commercial interactive spaces products and, probably, the future development of interactive learning spaces.

# Bibliography


Giovanni Agosta, Luca Borghese, Carlo Brandolese, Francesco Clasadonte, William Fornaciari, Franca Garzotto, Mirko Gelsomini, Matteo Grotto, Cristina Fra, Danny Noferi, et al. Playful Supervised Smart Spaces (P3S)–A Framework for Designing, Implementing and Deploying Multisensory Play Experiences for Children with Special Needs. In *Digital System Design (DSD), 2015 Euromicro Conference*, pages 158–164. IEEE, 2015.

Leonardo Amico. La Stanza Logo-Motoria. Un Ambiente multimodale interattivo per l'insegnamento a bambini in situazione di multi-disabilità. Master's thesis, Dipartimento di Ingegneria dell'Informazione, Università di Padova, 2012.

Adam Baimel, Rachel L. Severson, Andrew S. Baron, and Susan AJ Birch. Enhancing "theory of mind" through behavioral synchrony. *Frontiers in psychology*, 6, 2015.

R. Behringer and J. Elliott. Linking Physical Space with the Riemann Tonnetz for Exploration of Western Tonality. *Music Education*, 2009.

Victoria Bellotti, Maribeth Back, W. Keith Edwards, Rebecca E. Grinter, Austin Henderson, and Cristina Lopes. Making sense of sensing systems: five questions for designers and researchers. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 415–422. ACM, 2002.

David Benyon. Presence in blended spaces. *Interacting with Computers*, 24(4):219–226, Jul. 2012. doi: 10.1016/j.intcom.2012.04.005. URL `http://dx.doi.org/10.1016/j.intcom.2012.04.005`.

Tony Bergstrom, Karrie Karahalios, and John C Hart. Isochords: visualizing structure in music. In *Proceedings of Graphics Interface 2007*, pages 297–304. ACM, 2007.

Jonathan W. Bernard. Voice leading as a spatial function in the music of Ligeti. *Music Analysis*, 13(2/3):227–253, 1994.

Jamshed Bharucha and Carol L. Krumhansl. The representation of harmonic structure in music: Hierarchies of stability as a function of context. *Cognition*, 13(1):63–102, 1983.

M. Bick and T-F. Kummer. Ambient intelligence and ubiquitous computing. In *Handbook on Information Technologies for Education and Training*, pages 79–100. Springer, 2008.

Louis Bigo, Jérémie Garcia, Antoine Spicher, Wendy E. Mackay, et al. Papertonnetz: Music composition with interactive paper. In *Sound and Music Computing*, 2012.

Mark Billinghurst and Bill Buxton. Gesture based interaction. *Haptic input*, 24, 2011.

Sara A. Bly, Steve R. Harrison, and Susan Irwin. Media spaces: Bringing people together in a video, audio, and computing environment. *Commun. ACM*, 36(1):28–46, January 1993. ISSN 0001-0782. doi: 10.1145/151233.151235. URL `http://doi.acm.org/10.1145/151233.151235`.

Aaron F. Bobick, Stephen S. Intille, James W. Davis, Freedom Baird, Claudio S. Pinhanez, Lee W. Campbell, Yuri A. Ivanov, Arjan Schütte, and Andrew Wilson. The kidsroom: A perceptually-based interactive and immersive story environment. *Presence: Teleoperators and Virtual Environments*, 8(4):369–393, 1999.

Jürgen Bohn, Vlad Coroamă, Marc Langheinrich, Friedemann Mattern, and Michael Rohs. Social, economic, and ethical implications of ambient intelligence and ubiquitous computing. In *Ambient intelligence*, pages 5–29. Springer, 2005.

Richard A. Bolt. Spatial data management system. Technical report, DTIC Document, 1978.

Jan Borchers, Eric Lee, Wolfgang Samminger, and Max Mühlhäuser. Personal orchestra: a real-time audio/video system for interactive conducting. *Multimedia Systems*, 9(5):458–465, 2004.

Alan Bränzel, Christian Holz, Daniel Hoffmann, Dominik Schmidt, Marius Knaust, Patrick Lühne, René Meusel, Stephan Richter, and Patrick Baudisch. Gravityspace: tracking users and their poses in a smart room using a pressure-sensing floor. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 725–734. ACM, 2013.

Albert S. Bregman. *Auditory scene analysis: The perceptual organization of sound*. MIT press, 1994.

Candace Brower. A cognitive theory of musical meaning. *Journal of Music Theory*, pages 323–379, 2000.

William Buxton. Living in augmented reality: Ubiquitous media and reactive environments. *Video Mediated Communication. Hillsdale, NJ: Erlbaum*, pages 363–384, 1997.

Claude Cadoz, Marcelo M. Wanderley, et al. Gesture-music. In *Trends in gestural control of music*, volume 12, page 101. Paris: Ircam, 2000.

A. Camurri, S. Canazza, C. Canepa, G. L. Foresti, A. Rodà, G. Volpe, and S. Zanolla. The stanza logo-motoria: an interactive environment for learning and communication. In *Proceedings of SMC Conference 2010, Barcelona*, 2010.

Antonio Camurri, Barbara Mazzarino, Matteo Ricchetti, Renee Timmers, and Gualtiero Volpe. Multimodal analysis of expressive gesture in music and dance performances. In *International Gesture Workshop*, pages 20–39. Springer, 2003.

Antonio Camurri, Barbara Mazzarino, and Gualtiero Volpe. Analysis of expressive gesture: The eyesweb expressive gesture processing library. In *Gesture-Based Communication in Human-Computer Interaction*, pages 460–467. Springer Science, 2004.

Antonio Camurri, Corrado Canepa, and Gualtiero Volpe. 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*, pages 56–61. ACM, 2007.

Antonio Camurri, Corrado Canepa, Paolo Coletta, Barbara Mazzarino, and Gualtiero Volpe. Mappe per affetti erranti: a multimodal system for social active listening and expressive performance. In *Proceedings of the 8th Intl. Conference on New Interfaces for Musical Expression (NIME'08)*, pages 134–139, 2008.

J. C. Chung and C. K. Lai. Snoezelen for dementia. *Reviews.*, doi:10.1002/14651858.cd003152, 2002.

Andy Clark. *Being there: Putting brain, body, and world together again.* MIT press, 1998.

Martin Clayton, Rebecca Sager, and Udo Will. In time with the music: the concept of entrainment and its significance for ethnomusicology. In *European meetings in ethnomusicology.*, volume 11, pages 1–82. Romanian Society for Ethnomusicology, 2005.

Kathleen A. Corrigall and Laurel J. Trainor. Musical enculturation in preschool children: Acquisition of key and harmonic knowledge. *Music Perception*, 28(2):195–200, 2010.

Yvonne AW. De Kort and Wijnand A. Ijsselsteijn. People, places, and play: player experience in a socio-spatial context. *Computers in Entertainment (CIE)*, 6(2):18, 2008.

W. Jay Dowling, Kitty Mei-Tak Lung, and Susan Herrbold. Aiming attention in pitch and time in the perception of interleaved melodies. *Perception & Psychophysics*, 41(6):642–656, 1987. ISSN 1532-5962. doi: 10.3758/BF03210496. URL `http://dx.doi.org/10.3758/BF03210496`.

Douglas C. Engelbart. Nls teleconferencing features: The journal, and shared-screen telephoning. *Fall COMPCON 75 Digest of Papers*, pages 173–177, 1975.

Douglas C. Engelbart. Augmenting human intellect: a conceptual framework (1962). *PACKER, Randall and JORDAN, Ken. Multimedia. From Wagner to Virtual Reality. New York: WW Norton & Company*, pages 64–90, 2001.

Noel Enyedy, Joshua A. Danish, Girlie Delacruz, and Melissa Kumar. Learning physics through play in an augmented reality environment. *International Journal of Computer-Supported Collaborative Learning*, 7(3):347–378, 2012.

Noel Enyedy, Joshua A. Danish, and David DeLiema. Constructing liminal blends in a collaborative augmented-reality learning environment. *Intern. J. Comput.-Support. Collab. Learn.*, 10 (1):7–34, feb 2015. doi: 10.1007/s11412-015-9207-1. URL `http://dx.doi.org/10.1007/s11412-015-9207-1`.

Cumhur Erkut, Antti Jylhä, and Davide Rocchesso. 17 Heigh Ho: Rhythmicity in Sonic Interaction. *Sonic Interaction Design*, page 341, 2013.

Gilles Fauconnier and Mark Turner. *The way we think: Conceptual blending and the mind's hidden complexities*. Basic Books, 2008.

Jerry A. Fodor. *The modularity of mind: An essay on faculty psychology*. MIT press, 1983.

Anders Friberg and Johan Sundberg. Does music performance allude to locomotion? A model of final ritardandi derived from measurements of stopping runners. *The Journal of the Acoustical Society of America*, 105(3):1469–1484, 1999.

Masataka Goto. Smartmusickiosk: Music listening station with chorus-search function. In *Proceedings of the 16th annual ACM symposium on User interface software and technology*, pages 31–40. ACM, 2003.

Masataka Goto. Active music listening interfaces based on signal processing. In *2007 IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP'07*, volume 4, pages IV–1441. IEEE, 2007.

Masataka Goto, Kazuyoshi Yoshii, Hiromasa Fujihara, Matthias Mauch, and Tomoyasu Nakano. Songle: A web service for active music listening improved by user contributions. In *ISMIR*, pages 311–316, 2011.

Niall Griffith and Mikael Fernström. Litefoot: A floor space for recording dance and controlling media. In *Proceedings of the 1998 International Computer Music Conference*, pages 475–481, 1998.

Mark Grimshaw and Gareth Schott. Situating gaming as a sonic experience: the acoustic ecology of first person shooters. In *Digital Games Research Association (DiGRA)*, 2007.

Kaj Grønbæk, Ole Sejer Iversen, Karen Johanne Kortbek, Kaspar Rosengreen Nielsen, and Louise Aagaard. Interactive floor support for kinesthetic interaction in children learning environments. In *Human-Computer Interaction–INTERACT 2007*, pages 361–375. Springer, 2007.

Simone Gumtau, Paul Newland, Chris Creed, and Simon Kunath. Mediate: A responsive environment designed for children with autism. In *Proceedings of the 2005 International Conference on Accessible Design in the Digital World*, Accessible Design'05, pages 14–14, Swinton, UK, UK, 2005. British Computer Society. URL `http://dl.acm.org/citation.cfm?id=2227734.2227754`.

Masatoshi Hamanaka and Seunghee Lee. Sound scope headphones: Controlling an audio mixer through natural movement. In *Proceedings of the International Computer Music Conference*, pages 155–158, 2007.

Thomas W. Hedges and Andrew P. McPherson. 3D Gestural Interaction with Harmonic Pitch Space. *Proceedings of ICMC-SMC Conference*, 2013.

Simon Holland. *Learning about harmony with Harmony Space: an overview*. Springer 1994, 1994.

Simon Holland, Paul Marshall, Jon Bird, and al. Running up Blueberry Hill: Prototyping whole body interaction in harmony space. *Proceedings of the 3rd international Conference on Tangible and Embedded interaction*, 2009.

Eva Hornecker and Jacob Buur. 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*, pages 437–446. ACM, 2006.

Johanna Hoysniemi. International survey on the dance dance revolution game. *Comput. Entertain.*, 4(2), April 2006. ISSN 1544-3574. doi: 10.1145/1129006.1129019. URL `http://doi.acm.org/10.1145/1129006.1129019`.

Mu Hsen Hsu, Timothy K. Shih, and Jen Shiun Chiang. Real-time finger tracking for virtual instruments. In *Ubi-Media Computing and Workshops (UMEDIA), 2014 7th International Conference on*, pages 133–138. IEEE, 2014.

Lee Humphries. What to think about when you conduct: perception, language, and musical communication. *ThinkingApplied.com*, 1968.

Horace HS Ip, Ken CK Law, and Belton Kwong. Cyber composer: Hand gesture-driven intelligent music composition and generation. In *Multimedia Modelling Conference, 2005. MMM 2005. Proceedings of the 11th International*, pages 46–52. IEEE, 2005.

Ole Sejer Iversen, Karen Johanne Kortbek, Kaspar Rosengreen Nielsen, and Louise Aagaard. Stepstone: an interactive floor application for hearing impaired children with a cochlear implant. In *Proceedings of the 6th international conference on Interaction design and children*, pages 117–124. ACM, 2007.

S. Jabri, Z. Duric, H. Wechsler, and A. Rosenfeld. Detection and location of people in video images using adaptive fusion of color and edge information. In *In Pattern Recognition, Proc. of the 15th International Conference on*, volume 4, pages 627–630, 2000.

Robert JK. Jacob, Audrey Girouard, Leanne M. Hirshfield, Michael S. Horn, Orit Shaer, Erin Treacy Solovey, and Jamie Zigelbaum. Reality-based interaction: a framework for post-wimp interfaces. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 201–210. ACM, 2008.

Jan Jarvlepp. Pitch and Texture Analysis of Ligeti's Lux Aeterna, January 1982. URL http://www.ex-tempore.org/jarvlepp/jarvlepp.htm.

Kristoffer Jensen. Pitch gestures in generative modeling of music. In *Proceedings of the 7th International Conference on Exploring Music Contents*, CMMR'10, pages 51–59, Berlin, Heidelberg, 2011. Springer-Verlag. ISBN 978-3-642-23125-4. URL http://dl.acm.org/citation.cfm?id=2040789.2040794.

Hans-Christian Jetter, Florian Geyer, Tobias Schwarz, and Harald Reiterer. Blended Interaction–Toward a Framework for the Design of Interactive Spaces. In *Workshop DCIS*, volume 12, 2012.

Hans-Christian Jetter, Harald Reiterer, and Florian Geyer. Blended interaction: understanding natural human–computer interaction in post-wimp interactive spaces. *Personal and Ubiquitous Computing*, 18(5):1139–1158, 2014.

Mina C. Johnson-Glenberg, David Birchfield, Philippos Savvides, and Colleen Megowan-Romanowicz. Semi-virtual embodied learning-real world stem assessment. In *Serious Educational Game Assessment*, pages 241–257. Springer, 2011.

Mari R. Jones and Marilyn Boltz. Dynamic attending and responses to time. *Psychological review*, 96(3):459, 1989.

Sergi Jordà, Günter Geiger, Marcos Alonso, and Martin Kaltenbrunner. 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*, pages 139–146. ACM, 2007.

Patrik N. Juslin and Petri Laukka. Expression, perception, and induction of musical emotions: A review and a questionnaire study of everyday listening. *Journal of New Music Research*, 33 (3):217–238, 2004.

J. Juul. *Half-Real: Video Games Between Real Rules and Fictional Worlds*. MIT Press, 2011. ISBN 9780262516518. URL https://books.google.it/books?id=mZP6AQAAQBAJ.

Gunild Keetman and Carl Orff. *Music for Children*, volume Voll. 1 -5 of *Orff-Schulwerk - Music for Children*. Schott Music Ltd., London, score for voice and/or instruments edition, 1950 - 1954.

Peter Krogh, Martin Ludvigsen, and Andreas Lykke-Olesen. " Help Me Pull That Cursor" A Collaborative Interactive Floor Enhancing Community Interaction. *Australasian Journal of Information Systems*, 11(2), 2004.

Myron W. Krueger. Responsive environments. In *Proceedings of the June 13-16, 1977, national computer conference*, pages 423–433. ACM, 1977.

Myron W. Krueger. Environmental technology: making the real world virtual. *Communications of the ACM*, 36(7):36–37, 1993.

Myron W. Krueger, Thomas Gionfriddo, and Katrin Hinrichsen. Videoplace—an artificial reality. In *ACM SIGCHI Bulletin*, volume 16, pages 35–40. ACM, 1985.

George Lakoff and Mark Johnson. *Metaphors we live by*. University of Chicago press, 2008.

Marc Leman. *Embodied music cognition and mediation technology*. Mit Press, 2008.

Marc Leman, Michiel Demey, Micheline Lesaffre, Leon Van Noorden, and Dirk Moelants. Concepts, Technology, and Assessment of the Social Music Game" Sync-in-Team". In *Computational Science and Engineering, 2009. CSE'09. International Conference on*, volume 4, pages 837–842. IEEE, 2009.

Fred Lerdahl and Ray Jackendoff. *A generative theory of tonal music*. MIT Press, Cambridge, 1985.

Henrik Hautop Lund, Thomas Klitbo, and Carsten Jessen. Playware technology for physically activating play. *Artificial life and Robotics*, 9(4):165–174, 2005.

Pattie Maes. Artificial life meets entertainment: lifelike autonomous agents. *Communications of the ACM*, 38(11):108–114, 1995.

Pieter-Jan Maes, Denis Amelynck, Micheline Lesaffre, Marc Leman, and DK Arvind. The "conducting master": an interactive, real-time gesture monitoring system based on spatiotemporal motion templates. *International Journal of Human-Computer Interaction*, 29(7):471–487, 2013.

Prodromos Makris, Dimitrios N. Skoutas, and Charalabos Skianis. A survey on context-aware mobile and wireless networking: on networking and computing environments' integration. *Communications Surveys & Tutorials, IEEE*, 15(1):362–386, 2013.

Bill Manaris, David Johnson, and Yiorgos Vassilandonakis. Harmonic navigator: A gesture-driven, corpus-based approach to music analysis, composition, and performance. In *Ninth Artificial Intelligence and Interactive Digital Entertainment Conference*, 2013.

Teresa Marrin and Joseph Paradiso. The digital baton: a versatile performance instrument. In *Proceedings of the International Computer Music Conference*, pages 313–316, 1997.

Teresa Marrin and Rosalind Picard. The 'conductor's jacket': A device for recording expressive musical gestures. In *Proc. International Computer Music Conference*, pages 215–219, 1998.

Max V. Mathews. The radio baton and conductor program, or: Pitch, the most important and least expressive part of music. *Computer Music Journal*, 15(4):37–46, 1991. ISSN 01489267, 15315169. URL `http://www.jstor.org/stable/3681070`.

mDecks Music. Mapping tonal harmony, November 2012. URL http://mdecks.com/mapharmony.html.

D. R. Montello. *International Encyclopedia of the Social and Behavioral Sciences*. Oxford, Pergamon Press, 2001.

Axel G. E. Mulder. Towards a choice of gestural constraints for instrumental performers. *Trends in Gestural Control of Music*, pages 315–335, 2000.

Matteo Munaro, Filippo Basso, and Emanuele Menegatti. Openptrack: Open source multi-camera calibration and people tracking for rgb-d camera networks. *Robotics and Autonomous Systems*, 75, Part B:525–538, 1 2016. doi: http://dx.doi.org/10.1016/j.robot.2015.10.004. URL http://www.sciencedirect.com/science/article/pii/S0921889015002304.

Teresa Marrin Nakra, Yuri Ivanov, Paris Smaragdis, and Christopher Ault. The ubs virtual maestro: an interactive conducting system. In *NIME*, pages 250–255, 2009.

F. Pachet. Music listening: What is in the air? Technical report, Sony Computer Science Laboratory internal Report, 1999.

François Pachet, Olivier Delerue, and Peter Hanappe. Dynamic audio mixing. In *Proceedings of ICMC*, 2000.

Haris Papagiannakis, Margherita Antona, Stavroula Ntoa, and Constantine Stephanidis. A multimodal ambient intelligence environment for playful learning. *J. UCS*, 19(17):2617–2636, 2013.

Joseph Paradiso, Craig Abler, Kai-yuh Hsiao, and Matthew Reynolds. The magic carpet: physical sensing for immersive environments. In *CHI'97 Extended Abstracts on Human Factors in Computing Systems*, pages 277–278. ACM, 1997.

Joseph A. Paradiso and Neil Gershenfeld. Musical Applications of Electric Field Sensing. *Computer Music Journal*, (April):1–28, 1996.

Marianne Graves Petersen. Interactive spaces: towards a better everyday? *interactions*, 12(4): 44–45, 2005.

Jessica Phillips-Silver, C. Athena Aktipis, and Gregory A. Bryant. The ecology of entrainment: Foundations of coordinated rhythmic movement. *Music Perception: An Interdisciplinary Journal*, 28(1):3–14, 2010.

Jean Piaget, Margaret Cook, and W.W. Norton. *The origins of intelligence in children*, volume 8. International Universities Press New York, 1952.

Martin Pichlmair and Fares Kayali. Levels of sound: On the principles of interactivity in music video games. In *Proceedings of the Digital Games Research Association 2007 Conference" Situated play*. Citeseer, 2007.

Bogdan Pietraszewski, Slawomir Winiarski, and Sebastian Jaroszczuk. Three-dimensional human gait pattern–reference data for normal men. *Acta of Bioengineering and Biomechanics*, 14(3):9–16, 2012.

Dirk-Jan Povel and Erik Jansen. Harmonic factors in the perception of tonal melodies. *Music Perception*, 20(1):51–85, 2002.

Arthur S. Reber. Implicit learning and tacit knowledge. *Journal of experimental psychology: General*, 118(3):219, 1989.

Bruce Richardson, Krispin Leydon, Mikael Fernstrom, and Joseph A. Paradiso. Z-tiles: building blocks for modular, pressure-sensing floorspaces. In *CHI'04 extended abstracts on Human factors in computing systems*, pages 1529–1532. ACM, 2004.

Martin Rohrmeier and Patrick Rebuschat. Implicit learning and acquisition of music. *Topics in cognitive science*, 4(4):525–553, 2012.

Joseph Rovan and Vincent Hayward. Typology of tactile sounds and their synthesis in gesture-driven computer music performance. *Trends in gestural control of music*, pages 297–320, 2000.

M. Rudolf and M. Stern. *The grammar of conducting: a comprehensive guide to baton technique and interpretation*. [David Lewis Crosby collection]. Schirmer Books, 1994. ISBN 9780028722214. URL `http://books.google.it/books?id=qr8YAQAAIAAJ`.

Teppo Särkämö, Mari Tervaniemi, Sari Laitinen, Anita Forsblom, Seppo Soinila, Mikko Mikkonen, Taina Autti, Heli M. Silvennoinen, Jaakko Erkkilä, Matti Laine, et al. Music listening enhances cognitive recovery and mood after middle cerebral artery stroke. *Brain*, 131(3): 866–876, 2008.

Donald A. Schön. The theory of inquiry: Dewey's legacy to education. *Curriculum Inquiry*, 22 (2):119–139, 1992. doi: 10.1080/03626784.1992.11076093. URL `http://dx.doi.org/10.1080/03626784.1992.11076093`.

Sertan Şentürk, Sang Won Lee, Avinash Sastry, Anosh Daruwalla, and G. Weinberg. Crossole: A gestural interface for composition, improvisation and performance using kinect. In G. Essl, B. Gillespie, M. Gurevich, and S. O'Modhrain, editors, *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME)*, pages 449–502, Ann Arbor, Michigan, May 21-23 2012. University of Michigan, University of Michigan.

John A. Sloboda. *The musical mind: The cognitive psychology of music*. Oxford University Press, 1985.

Sundeep Teki, Maria Chait, Sukhbinder Kumar, Katharina von Kriegstein, and Timothy D. Griffiths. Brain bases for auditory stimulus-driven figure–ground segregation. *The Journal of Neuroscience*, 31(1):164–171, 2011.

Florian Thalmann and Guerino Mazzola. The bigbang rubette: Gestural music composition with rubato composer. In *Proceedings of the 2008 International Computer Music Conference*, 2008.

Michael H. Thaut, Gerald C. McIntosh, and Volker Hoemberg. Neurobiological foundations of neurologic music therapy: rhythmic entrainment and the motor system. *Frontiers in psychology*, 5, 2014.

Barbara Tillmann, Jamshed J. Bharucha, and Emmanuel Bigand. Implicit learning of tonality: a self-organizing approach. *Psychological review*, 107(4):885, 2000.

Leng-Wee Toh, Wilber Chao, and Yi-Shin Chen. An interactive conducting system using kinect. In *Multimedia and Expo (ICME), 2013 IEEE International Conference on*, pages 1–6. IEEE, 2013.

Andries van Dam. Post-wimp user interfaces. *Commun. ACM*, 40(2):63–67, February 1997. ISSN 0001-0782. doi: 10.1145/253671.253708. URL `http://doi.acm.org/10.1145/253671.253708`.

Giovanna Varni, Gaël Dubus, Sami Oksanen, Gualtiero Volpe, Marco Fabiani, Roberto Bresin, Jari Kleimola, Vesa Välimäki, and Antonio Camurri. Interactive sonification of synchronisation of motoric behaviour in social active listening to music with mobile devices. *Journal on Multimodal User Interfaces*, 5(3-4):157–173, 2012.

Bill Verplank. Interaction design sketchbook, 2003. URL `https://ccrma.stanford.edu/courses/250a/lectures/`.

Luc Vincent. Morphological area openings and closings for grey-scale images. In Ying-Lie O, Alexander Toet, David Foster, HenkJ.A.M. Heijmans, and Peter Meer, editors, *Shape in Picture*, volume 126 of *NATO ASI Series*, pages 197–208. Springer Berlin Heidelberg, Berlin, 1994.

Yon Visell, Severin Smith, Alvin Law, Rishi Rajalingham, and Jeremy R Cooperstock. Contact sensing and interaction techniques for a distributed, multimodal floor display. In *3D User Interfaces (3DUI), 2010 IEEE Symposium on*, pages 75–78. IEEE, 2010.

Gualtiero Volpe and Antonio Camurri. A system for embodied social active listening to sound and music content. *Journal on Computing and Cultural Heritage (JOCCH)*, 4(1):2, 2011.

Peter Vorderer and Jennings Bryant. *Playing video games: Motives, responses, and consequences*. Routledge, 2012.

Mark Weiser. The computer for the 21st century. *Scientific american*, 265(3):94–104, 1991.

Daniel Wigdor, Chia Shen, Clifton Forlines, and Ravin Balakrishnan. Table-centric interactive spaces for real-time collaboration. In *Proceedings of the working conference on Advanced visual interfaces*, pages 103–107. ACM, 2006.

Matthew Wright. Open sound control-a new protocol for communicationg with sound synthesizers. In *Proceedings of the 1997 International Computer Music Conference*, pages 101–104, 1997.

Tomoyoshi Yoshida, Shoichi Takeda, and Sayoko Yamamoto. The application of entrainment to musical ensembles, 2002.

S. Zanolla, S. Canazza, A. Rodà, A. Camurri, and G. Volpe. Entertaining listening by means of the Stanza Logo-Motoria: an Interactive Multimodal Environment. *Entertainment Computing*, 2013(4):213–220, 2013.