Studies on customisation-driven digital music instruments

Bruno Zamborlin

October 2014

Thesis jointly submitted to
Goldsmiths, University of London
and Université Pierre et Marie Curie - Paris 6, École Doctorale EDITE
for the degree of Doctor of Philosophy.
Supervised by:
Marco Gillies, Frédéric Bevilacqua, Mark d'Inverno, Gérard Assayag

Dedicated to Camila.

Declaration

This thesis is a presentation of my original research work. Whatever contributions of others are involved, every effort is made to indicate this clearly, with due reference to the literature, and acknowledgement of collaborative research and discussion.

In my capacity as supervisor of the candidate's thesis, I certify that the above statements are true to the best of my knowledge.

Date:

Abstract

From John Cage's Prepared Piano to the turntable, the history of musical instruments is scattered with examples of musicians who deeply customised their instruments to fit personal artistic objectives, objectives that differed from the ones the instruments have been designed for. In their digital counterpart however, musical instruments are often presented in the form of closed, finalised systems with a-priori symbolic rules set by their designer that leave very little room for the artists to customise the technologies for their unique art practices; in these cases the only possibility to change the mode of interaction with digital instrument is to reprogram them, a possibility available to programmers but not to musicians.

This thesis presents two digital music instruments designed with the explicit goal of being highly customisable by musicians and to provide different modes of interactions, whist keeping simplicity and immediateness of use. The first one leverages real-time gesture recognition to provide continuous feedback to users as guidance in defining the behaviour of the system and the gestures it recognises. The second one is a novel tangible user interface which allows to transform everyday objects into expressive digital music instruments and whose sound generated strongly depends by the particular nature of the physical object selected.

Résumé

Cette thèse explore un nouveau paradigme d'interaction pour le design d'instruments de musique numérique, en considrant comme fondamentaux les processus de customisation. Plutôt que de demander aux musiciens d'apprendre des règles spécifiques imposées par les instruments, j'explore des techniques d'adaptations pour faciliter la matrise des instruments par les musiciens, reflétant leur style et devenant une partie intégrante de leur production artistique. Deux nouveaux systèmes numériques sont réalisés avec un objectif commun: faciliter le phénomène de customisation dans le processus d'apprentissage. Le premier travail présenté est un système de reconnaissance de geste qui guide les utilisateurs dans la définition de leurs propres vocabulaires gestuels. Ce système leur permet de les utiliser pour contrôler les médias numériques en s'adaptant á leurs pratiques artistiques. Le deuxième travail est un instrument mobile basé sur des transducteurs piézoélectriques qui permettent aux utilisateurs de transformer facilement des objets physiques quotidiens en des instruments de musique originaux. Ce système convertit les vibrations créées lorsque les utilisateurs touchent ou frappent les objets en son musical. Plus précisément, le système met en oeuvre des techniques de synthèse inspirés aux modèls physiques qui dépendent des propriétés acoustiques de l'objet lui-même et de la façon dans la quelle il est touché par l'utilisateur. Cette approche á la musique 'reality-based' est évaluée par une série d'études d'utilisation dans des situations concrètes du monde réel.

'My topic is the shift from 'architect' to 'gardener', where 'architect' stands for 'someone who carries a full picture of the work before it is made', to 'gardener' standing for 'someone who plants seeds and waits to see exactly what will come up"

- Brian Eno

Contents

List of Figures

Chapter 1

Introduction

By utilising our body movements, gestural digital interfaces are powerful tools to provide accessible and expressive ways to explore digital media. These properties establish an intriguing approach for the design of new Digital Music Instruments (DMI) in which high degree of expressivity and intimacy with the technology is required.

Although every instrument can be customised to a certain extend, very few instruments are designed with the explicit purpose of encouraging such a process. If the sound of digital instruments has always been highly customised by musicians and sound designers, for long time keyboards and knobs have been the only way to control them and alternative interfaces remained less explored. Since 2000, a growing number of researchers and artists have focused their work on the development of novel gestural interfaces, the NIME conference being one of the most prominent examples in this field. It is true that in the case of open-source software, coders can reprogram digital instruments. However this process requires technical expertise that makes it unachievable for most of the musicians (*"performers are not necessarily programmers"* [Jordà, 2004a]), and even for them this is often a difficult task that can distract from the main artistic objective.

In this thesis two different systems are presented. They both share the common goal of allowing users to customise the way of interacting with them in order to personalise them for specific artistic needs, without the need to reprogram them. At the same time, they aim to achieve what Perry Cook defined as *"In-*

stant music, subtlety later" [Cook, 2001] and David Wessel named *"Low entry fee with no ceiling to virtuosity"* [Wessel and Wright, 2002]; both systems are designed to be easy and immediate to use, whilst at the same time allowing the possibility for advanced uses.

The first system leverages real-time gesture recognition to provide continuous feedback to users as guidance in defining the behaviour of the system and the gestures it recognises. The second one is a novel tangible user interface which allows to transform everyday objects into expressive digital music instruments and whose sound generated strongly depends by the particular nature of the physical object selected.

1.1 Context

There has been a constant trend throughout the history of music performance in modifying acoustic musical instruments and using them in ways different from those they have been conceived for. Among well-known examples such as John Cage's 'Prepared Piano' or Robert Fripp's 'New Standard Tuning', the research for bespoke sounds and interaction techniques created unique performances, challenging instrument designers and sticking forever into the memory of spectators.

This phenomenon has been central in the research dedicated to DMI. Several digital instruments are designed and built by the composers themselves in order to characterise specific performances. Recent mainstream works such as Björk's 'Crystalline' and Imogen Heap's 'Me and the Machine' have been created respectively using the custom-made *gameleste* (a celesta modified with elements of gamelan and sensors) and the *Mi.Mu* musical gloves. This type of work popularises a large body of computer-music and avant-garde performances, to demonstrate how the desire for performing digital music in personal and surprising new ways is rapidly growing. With computers becoming faster and sensors becoming more accurate and affordable, 266 different instruments were presented in only the first 8 years of the conference on New Interfaces for Musical Expression (NIME) [Newton and Marshall, 2011]. This does not even include the ones developed outside the academic world. This desire for unique instruments and performative gestures highlights the importance of *customisable* DMIs, that moves the focus of the instrument from a tool for sound production to an integral part of the stage and the artistic work itself.

However, this possibility is generally reserved for players endowed with high technical skills. In order to adapt their digital instruments, players need to be expert in instrumental craftsmanship or, in the digital world, be able to understand and edit software. Even so, this process can require much time and experimentation, breaking the flow of the composition and distracting them towards technical developments rather than the artistic outcomes of the piece. One of the challenges is to provide artists and designers the capability of customisation of the interaction without interrupting or breaking the interaction with the instrument.

Marathe and Sundar [Marathe and Sundar, 2011] report that in (non musical) HCI literature, the ability of an application to be customised is directly proportional to engagement level for the following reasons:

- to make a system's goals fit their own ([Dyck et al., 2003]);
- to make the system more efficient ([Mackay, 1991] , [Page et al., 1996]);
- to manage complexity;
- to dictate outcome;
- to manage information overload.

Authors show how the customisation is directly associated not just with a sense of control (SOC) but also with a sense of identity (SOI) ("*a sense of conveying one's self-representation through use of technological tools*"), described as *ownership* by Alan Dix [Dix, 2007]. The study clearly shows that SOC is a consequence of SOI, and therefore having control is motivated by expressing our identity in the customised application. This is not true in the other sense, as identity is caused by many reasons and control is one of them. Even more, such a relationships varies deeply with users level of expertise: for non-expert users the development of a sense of identity proved to be even more important than the sense of control.

Magnusson states that *technology can never be neutral* [Magnusson, 2009]. Implicitly or explicitly, technology will always contain guidance for determined usages, affected by the cultural, social and technological environment where the technology was born. Music technology is no exception, and it is common in artistic practices to attempt to adapt technologies, whether by interpretation of usage or by physical adaptation.

Although a certain rigidity is required in an instrument design and can even be a source of inspiration, the task of specifying the exact rules of interactions can be shifted away from the designer and towards the performer. If an instrument is designed explicitly to incorporate possibilities for flexible customisation, then the interaction with the instrument will vary more from performer to performer and will reflect more their peculiar artistic objectives. If gestures and modes of interaction can be customised by the users themselves, they will correspond more to the metaphors and meaning that the users build in relation with their functions in the sound generation, supporting Paul Dourish observation that *"users and not designers manage coupling"* [Dourish, 2004].

In this work I present techniques that encourage users to adapt, design and shape the gestures and physical interfaces. These systems are designed to have *low entry fees*, being easy and immediate to use without the need to program and aimed for users with and without specific technical knowledge. The customisation of the gestures and interfaces happens during the interaction workflow itself, generating sound continuously over time with low latency in order to provide meaningful guidance to the users. This easy, program-free customisation with instant sound feedback allows performers to freely *explore* different gestures and modes of interaction, quickly understanding their relations with the sound generated. This work aims to provide users with a sense of ownership and identity with the customised instrument.

Let us consider as an example a full-body gesture recognition system employed for a dance piece. If the system can recognise a pre-defined set of standard gestures only, then dancers are forced to perform such gestures in the piece in order to use the system, and the choreography has to be adapted for such a system. On the other hand, if a dancer can personalise the system to recognise the movements of a specific piece, instead of having to perform standard encoded ones, then the system could be better adapted to the performers idiosyncrasies and to the choreography, acquiring a behaviour which is unique for that piece and, in a way, becoming part of the piece itself.

As a second example, let's consider the case of a percussionist who performs live with their own instrument and electronics. Physical instruments' aesthetics, as well as the gestures used to perform with them, are very important aspects in the presence of a live show and the possibility to personalise them is an appealing feature for a performance scenography. Enabling the possibility of easily explore the use of wider palettes of gestures and physical objects whilst preserving the possibility to fine tune the sound synthesis could enable the performer for bespoke and original live set ups.

These two examples are not coincidental. They cover very specific scenarios and represent very specific issues that can be covered in the two systems that are presented in this work, as described in the next section.

The goal of this thesis is to embed the customisation process of the gestures and the physical interfaces in the interaction workflow with the instrument and to consider it an integral part of its design whilst preserving the '*low entry fees*' paradigm. The two instruments presented in this work explore techniques that enable players to customise the way systems react to their actions or to change their physical and aesthetic properties without breaking the flow of their experience. They focus on providing constant and meaningful feedback as guidance for their users during both the customisation and the performance processes. This enables the performers to achieve better control of the system and smoothes the personalisation process to make it more pleasant and effective.

1.2 Proposed methodology

In this work, I will first review the existing literature. This can be subdivided into material dealing with embodied interaction and with literature on customisation and appropriation of technology in HCI. I will then transpose these perspectives and insights into the domain of music technology, highlighting commonalities and differences between these two domains.

Two different instruments which attempt to accomplish such properties will then be presented. These instruments implement different algorithms for the analysis of users input. To evaluate the *performances* of such algorithms, a quantitative analysis methodology is adopted, measuring their performance using a pertinent input dataset. To evaluate the *effectiveness* of these properties for accomplishing our general goals, several user-studies are presented, which happened both in controlled environments with a small number of participants and in real world situations such as live shows and primary schools. Data is collected by observing performers using the instruments in these situations, asking them to fill out questionnaires and by video-recording semi-structured interviews. I will present data obtained and perform analysis upon it in relation to the proposed properties and general objectives. Finally, a description of several artistic outcomes in which the author collaborated with artists using these systems will be presented. Materials from these will be offered to the reader as examples of usage in specific contexts.

1.2.1 Case studies

In this work the design, realisation and evaluation of two gestural DMI are presented. Both instruments, although through entirely different approaches, offer ways for their players to invent and customise the interaction they have with them, whilst keeping them simple and immediate to use.

The first system leverages an existing algorithm for realtime gesture recognition called Gesture Follower developed by Bevilacqua, et al. [Bevilacqua et al., 2007, 2010], which can be applied to almost any existing gestural interface which streams data to a computer. The features of the algorithm are applied for the realisation of GIDE, an acronym for *"Gesture Interaction DEsigner"*. GIDE is an end-user application where a novel set of visualisation and sonification techniques are presented in order to make the behaviour of the algorithm more transparent and easier to understand, as well as making the application immediate and pleasant to use. The realtime nature of the Gesture Follower algorithm is used to provide various forms of feedback which offer precise guidance to the users in the definition of their own gesture templates, or gesture vocabulary. The resulting system is a personalised gesture recognition tool which can be easily tailored by end-users for their own needs and artistic practice for the control of digital media. Three examples of audiovisual applications, a quantitative evaluation of the recognition algorithm employed and an user-studies with 23 participants are described in the chapter.

The second system presented in this work is a novel tangible acoustic interface called Mogees, which aims to convert physical everyday objects into DMIs by using a single piezo-transducer. This augmentation process is immediate and straight-forward and no calibration is required. Users place the transducer onto objects they want to augment and start playing with them. The sound synthesis is driven directly by the signal from the sensor, allowing for continuous and expressive control. A function called 'capture' enables users to automatically set the parameters of the synthesis by touching the object they want to play just once, encouraging the exploration of different setups. Users can trigger musical notes by generating vibrations of different frequencies through the gestures they perform. Both sound synthesis and gesture analysis happen automatically and very few parameters are exposed to the users. This highly constrained environment allows users to focus their attention on the physical interaction with the object to easily explore different combinations and forms of interactions, without having to handle calibration or reprogramming tasks. In this work I describe the techniques and design strategies adopted to accomplish these goals, as well as describing the analysis and synthesis techniques employed. The work is then evaluated both quantitatively, testing the system against a dataset, and qualitatively, through different user-studies and artistic works.

1.2.2 Structure of the work

The following six chapters provide the core of the work undertaken.

Chapter 2 establishes the literature review of this work. Existing literature about embodied interaction and customisation of technology is reviewed both in the broad Human-Computer Interaction domain and specifically in music technology. I then review gestural interfaces and the different analysis techniques that are relevant to achieving the goals presented in the previous section. The work moves then to the domain of Tangible Acoustic Interfaces, a particular branch of Tangible User Interfaces that relies on audio as control signal.

Chapter 3 presents GIDE, a highly customisable and usable realtime gesture recognition system. This chapter shows how continuous control *from* the user, combined with continuous feedback *to* the user, can enable a series of visualisation and signification techniques useful to provide guidance in making gesture-based systems easier to understand, to use and to personalise even by users not expert in computing. The process of making the personalisation process itself more pleasant and effective is also explored. The algorithm used by GIDE is evaluated quantitatively and its performances are compared to more standard approaches, which however do not satisfy our desired properties. A user-study is finally presented, observing 23 participants using this system and interviewing them. The study aims to evaluate and compare the different visual and auditory feedback techniques described in the chapter and their effectiveness in simplifying and guiding the users to interact with the system.

Chapter 4 presents a new Tangible Acoustic Interface called *Mogees*, which allows users to transform a wide range of physical everyday objects into DMIs. The algorithm used for analysis of users' input is described and evaluated against other techniques using a database of gestures. Again, to evaluate the practical effectiveness of the techniques employed in the system, a user study with 17 participants comprising musicians and non musicians is presented.

Chapter 5 discusses the results of an experiment employing Mogees by a teacher in a primary school in Birmingham with a class constituted by 18 pupils between 7 and 9 years old. This study provides an ideal context to evaluate how the system described in chapter 4 has been employed in a real-world context of music education. Qualitative and quantitative results are discussed, in order to give a perspective of the system from the eyes of the pupils and the teacher herself.

Chapter 6 presents a diametrically opposite point of view of the same system, that of the professional music composer Ed Handley, part of the electronic dance music duo Plaid. Plaid and I collaborated together for 18 months, designing different performances and productions, all centred upon the Mogees system. This *practice-based* research is discussed in the chapter through a series of open questions, describing the motivations of the composer in adopting the system for his work and analysing his workflow and how he adapted Mogees for his environment. Besides several performances and a musical video, this collaboration resulted in the usage of Mogees for the Plaid music album *'Reachy Prints'* [Plaid, 2014].

Chapter 7 presents the conclusions of this thesis. I draw here the connections between the original motivations of this work and its potential academic, industrial and artistic impact. I contextualise the development of this work with the personal journey undertaken during these years, discussing the advantages and drawbacks of the presented approach against more traditional ones, anticipating potential critical objections and suggesting insights for future works.

Chapter 2

Literature Review

This chapter builds the body of the literature review of this thesis. In section 2.1, the theoretical foundations behind embodied interaction and customisation are reviewed in both the musical domain and in the broader field of Human-Computer Interaction (HCI). The phenomenon of transformation of musical instruments is discussed, revealing the motivations of performers in the customisation practice. I then review existing literature about appropriation of technologies, both in its original HCI domain and in recent applications in the domain of music technology. In section 2.2, the world of gestural interfaces is introduced and techniques for the implementation of Digital Music Instruments are presented, highlighting their implications for the goals of our work. Concepts presented in this section establish the technical foundations for the system presented in chapter 3. Finally, section 2.3 focuses on Tangible Acoustic Interfaces, a particular family of tangible interfaces which use audio as control signal. This body of work establishes the technical starting point for the system presented in chapter 4.

2.1 On embodiment and customisation of technologies

This section starts by reviewing the foundations of *embodiment* in the disciplines of philosophy of the mind and consequentially in Human-Computer Interaction (2.1.1). I then transpose these concepts in the musical domain, highlighting its relevance in the interaction with DMIs. In 2.1.2, the phenomenon of *transformation* of musical instruments is reviewed respectively in their acoustic, digitally-augmented and purely digital correspondents. The motivations of this phenomenon are analysed and the challenges from a player point of view are pointed out.

However, in this work modification of technology is not only observed as a consequence of specific phenomenons of interactions, but also seen as a concrete design feature that can be encouraged with the achievement of specific interaction design strategies. Therefore, in section 2.1.3 I move towards a design prospective by presenting existing works about customisation, formalising such forms of interaction and its advantages. Finally, I review concrete guidelines proposed in the literature to design for customisation of technologies.

2.1.1 Embodied interaction

When musicians perform with a musical instrument, a certain degree of 'intimacy' has to be reached and players enter in a status of confidentiality with the instrument during which the focus of the attention is totally directed to the music produced and not to the interaction with the instrument itself. Different studies have been undertaken to better understand and define this 'intimacy', not just in the music domain but in the broader domain of interaction with technology.

A definition Paul Dourish, in his book "Foundations of Embodied Interaction" [Dourish, 2004], defines what he calls an *embodied interaction* as *the creation, manipulation and sharing of meaning through engaged interaction with artefacts.* By analysing two separate fields of HCI, tangible computing and social computing, he observes many common points in the way they both leverage on our natural skills to interact with the real world and our familiarity with it. Tangible computing focuses on our skills in interacting with physical objects, where social computing observes the way we interact with other people around us. What tangible and social computing have in common, for Dourish, is not only that they both exploit metaphors for interaction: they both study relationships between an action and its meaning. And this relationship is nothing less than the primary goal of phenomenology.

Active externalism In 1998 Andy Clark and David Chalmers introduced the concept of *active externalism* [Menary, 2010]. Their Extended Mind thesis rejects both the Cartesian dualism between body and mind (*what is outside the body is outside the mind*) and the radical externalism about meaning for which everything we experience is part of our mind. This approach defines the boundaries between mind, body and environment as a cognitive phenomenon rather than a physical one. Human cognitive process strongly relies on the support that the environment offers and the way humans can access and modify it. When we use our fingers as working memory for a mental calculation, fingers become part of the mental process of counting. As our fingers are always accessible and we know how to control them, we can establish a *reliable coupling* with them, perhaps in the same way we could with a pocket calculator if it was embedded somehow in our hands. Clark and Chalmers illustrate this concept through two fictional characters, Otto and Inga, who are both travelling to a museum. Otto is affected by Alzheimer's disease and wrote information about the address of the museum on a notebook, whereas Inga relies entirely in her memory. Inga can be thought to have had a belief as to the address of the museum before querying her memory, in the same way as Otto can be said to have held a belief on such location before consulting his notebook. The argument presented is that the only difference between Inga and Otto is that for the former the memory has been stored internally in the brain and for the latter the mind has been *extended* to the notebook. For Clark and Chalmers, Otto has been able to integrate the notebook into his mind because the notebook was constantly and immediately accessible and he was able to establish a reliable coupling with it. Portability, accessibility and predictability assume a big importance in this view as a fundamental requirement for coupling and embodiment.

Embodied musical instruments The theory of embodiment has been transposed to the musical world as well. Marc Leman's Embodied Music Cognition and Mediation Technology [Leman, 2008] argues for the importance of transparent technology that invisibly mediates between actions and perceptions, experienced as "behavioural resonance with sound energy". Topics of direct and indirect involvement with music are discussed and related with the theory of flow [Csikszentmihalyi, 1991]. This subject will be reconsidered when discussing the world of tangible acoustic interfaces later in the chapter (2.3).

With the exponential increasing of computer power and sensor technology becoming more and more accurate and affordable, the number of DMI is now increasing year after year. As already mentioned in 1, Newton and Marshall [Newton and Marshall, 2011] observed that only during the first 8 years of the conference on New Interfaces for Musical Expression (NIME), 266 different instruments have been developed but only very few of them have been use by a large audience.

The design of expressive and effective DMI is a challenging task in Computer-Human Interaction not only because the range of sounds that can be generated is immense, but also because, as the controller and the sound engine can also be dissociated one from the other, the relationship between one and the other is arbitrary. This point, as observed by Armstrong [Armstrong, 2006], can be one of the causes of a disconnection between the performer, the instrument and the audience due to a lack of understanding in the coupling between the performance and the generated sound.

Essl and O'Modhrain [Essl and O'Modhrain, 2006], when describing a DMI called PebbleBox, stress on the importance on *enaction*, defined as "the necessary and close link between action and perception", and how this closely relates with embodiment. Describing the interaction design around the PebbleBox, they observe the importance of such a relationship and how this depends on users expectations about the instrument and its affordances.

Intimacy and virtuosity Moore [Moore, 1988], discussing issues related to the adoption of the MIDI protocol, introduced the notion of *intimacy* with musical instruments. This concept has been developed further by Wessel and Wright [Wessel and Wright, 2002] which proposed general guidelines for the design of DMI emphasising the notion of intimacy between the performer and the instrument. They focus on the importance of ease to use of the instruments when approached by the performer for the first time (*low entry fee*), whereas also it should also provide *no ceiling to virtuosity* when higher degree of intimacy is reached. Wessel and Wright stress on the link between action and perception also in terms of timing, claiming that the *latency* between the gesture of the performer and the generated sound should be lower than 10 milliseconds and with extremely low variance (less then 1 millisecond) and *continuous* over time. Finally, authors argue for the importance of clear and simple strategies to program relationship between gesture and sound metaphors for music control. This last point is highly relevant for the two systems described in chapter 3 and 4, which describe two DMIs in which such a relationship is defined by users themselves.

Magnusson [Magnusson, 2009] observed how virtuosity in the digital domain is primarily concerned with the degree for which users master the cognitive associations between themselves and the technology, highlighting the emergence of high skilled computer musicians in designing and implementing their own custom instruments. This subject brings us directly to the next section of this review: the Art of Transforming Musical Instruments.

2.1.2 The art of transforming musical instruments

We can easily observe how the phenomenon of customise instruments has been regularly present throughout the history of music. Musicians often modify their instruments by changing their properties or by using them in personal ways for which the instruments were not originally conceived for. Customising an instrument is a phenomenon which mirrors the personality of the executor in the performance itself, reflecting a research for unique sounds and for unique aesthetics, as well to reach a higher degree of intimacy and embodiment with the instrument.

This section reviews this phenomenon starting from some historical landmarks and then moving to the realm of contemporary digital instruments.

Modifying acoustic instruments: historical landmarks The concept of adaptation in music is far from being new and has always been applied in acoustic instruments. Acoustic instruments have been adapted by composers to obtain sounds different from the ones they have been designed for. Many composers and performers physically modified almost any type of acoustic instrument by adding new components or employing them in ways other than the ones they have been built for.

In 1913 for his piece Piège de Méduse, Erik Satie instructed performers to place paper sheets on the strings of the piano to imitate the mechanical sound of a monkey puppet. 25 years later avant-garde composer John Cage, when working to the piece Bacchanale, added various resonant objects between the strings and the hammers of the piano so that a wide range of percussive sounds could be obtained by a single piano player. He successively referred to these techniques coining the term *prepared piano*, which became a common practice among contemporary and avant-garde performers.

Prepared guitars are another common example of the same kind. By placing different objects on top or between the strings of the guitar, or simply by tuning them in peculiar and unique ways, performers altered the timbre of the instruments to tailor it to their own compositions. Norwegian composer Bjørn Fongaard has perhaps been one of the first ones to adopt such technique, although the prepared guitar methodology has then been applied and formalised by many contemporary guitarists, among the ones English improvisers Keith Rowe and Fred Frith stand out, making their prepared guitars the emblems of their aesthetic and sound. This not to mention the number of drummers, percussionists and players of wind-instruments which modified their instruments to obtain bespoke sounds.

In 1939, John Cage adapted the usage of two turntables so as to modify their playback speed and pitch for his composition 'Imaginary Landscape No. 1'. Who knew that that particular modification of the technology, created for one particular music piece, would have prepared the terrain, 30 years later in the Bronx, for artists such as DJ Kool Herc, Grandmaster Flash and Afrika Bambaataa that pioneered the use the turnable as a musical instrument inventing the turntablism [Hansen, 2014].

Digital augmentations The *preparation* practice has more recently involved digital tools as well. By adding sensors to acoustic instruments or on the body of the performers, software programs can analyse information about the performers and generate digital audio which is coupled with the acoustic signal generated by the instrument. In 1986 Tod Machover and MIT Media Lab developed the concept of *hyperinstruments*, in which acoustic instruments such as violin or piano are used as input interface for computer-music software. Another example is the IRCAM Augmented Violin [Bevilacqua et al., 2006], in which bow acceleration is analysed in real-time to characterise bow styles and recognise pre-recorded patterns. These instruments are often referred in the literature as *augmented instruments*.

In 2011 Newton et al. [Newton and Marshall, 2011] described a users-test in which 10 participants were guided in the process of augmenting their own instruments by themselves, using Phidgets (www.phidgets.com) sensors and MaxMSP patches. By selecting the acoustic instrument, the sensors and the relation (*mapping*) with the sound synthesis, the musicians, although guided in the experiment, were authors of the augmentation process and the final instrument created typically reflected musicians personal artistic tastes.

McPherson et al. [McPherson et al., 2013] presented a system called TouchKeys which, by integrating capacitive touch sensors on an existing keyboard, offers the possibility to track players' fingers positions continuously over time in order to control parameters such as vibrato and pitch bending. The challenge of this type of systems is to enhance the expressivity bandwidth of the keyboard without compromising the existing expertise of the keyboard players, hardly acquired over the years. Guidelines for designers of augmented-instruments are provided, pointing out the importance that the interaction with the additional sensors must have minimal interference with the movements required to perform the original instruments, in order to avoid to engage the augmented features unintentionally. Movements data-log is suggested in order to find patterns that are not part of the traditional use of the instrument.

In [McPherson and Kim, 2012], the authors present a user-study where 6 composers are asked to compose a piece using the magnetic resonator piano. The study focuses on relaxing the constraints of the instrument and in observing the different degrees of appropriation that the composers adopted when using the instrument over an entire year. The study demonstrates usages of the instrument never expected by the authors, that constitute an invaluable guidance in the design of the instrument itself (for example, in the usage of different type of notations to write the piece or in leveraging existing piano techniques at very different levels). The paper concludes with a guideline for user-centred DMI design, focusing on the importance of long-term relationships between designers and players and on connecting with familiar models in order to use players pre-existing music skills.

Programming tools As observed at the beginning of this thesis, often the only way to deeply customise digital musical instruments is the one of reprogramming them. The last two decades brought to life a variety of different software environments for the creation of audio software that shares the common goal of allowing expert users to develop their own audio tools and instruments. Often, the digital instruments created using these tools allow users to modify them and adapt them to their own needs. This thesis strongly supports Jorda ([Jordà, 2004a]) and Cook ([Cook, 2001]) positions that musicians should not be required to know computer languages in order to use (and customise) their instruments. However, the birth of so many music-oriented computer languages clearly proves a strong interest for personal ad-hoc DMI and it is therefore worth to be mentioned.

Perhaps the most successful of these environments are PureData (http://puredata.info/) and Max (http://http://cycling74.com/products/max/), which allow for software programming by connecting together different pre-compiled software components visually represented by boxes.

This family of languages made much easier for electronic musicians to create new instruments or download the code that other musicians shared (sometime referred as 'patch'), modify them for specific needs and then share them again. This phenomenon gave birth to entire dedicated communities of specialised programmers and the outcome of such works is very often not limited to the music generated by the electronic instruments, but also by the instruments themselves and the technical and social practice that brought to their creation.

Conclusions Different examples of transformation of acoustic, digitally-augmented and entirely digital instruments have been reviewed. Modifying musical instruments is a common desire for many types of musicians and performers from different backgrounds. However, these modification processes often require a great mastery in audio technology and programming skills and are not achievable from the majority of musicians. Moreover, these modifications often focus on the sound rather than on the interfaces themselves.The next section highlights some practical existing guidelines in the broader context of Human-Computer Interaction that relate directly with this subject.

2.1.3 Designing for appropriation

Definitions A concept similar, but distinct, to the one of customisation of technologies has been studied in the world of Human-Computer Interaction: the one of appropriation.

For Paul Dourish [Dourish, 2004] (pages 172, 204-205), appropriation concerns *the way in which practices and technologies evolve around each other*, it is *the process by which people adopt and adapt technologies, fitting them into their working practices.*

> *"Users play a much more active role in determining precisely how a technology will meet their needs - needs that are continually changing, and that will be satisfied using a variety of features of the setting, of which the technological artifact is only one"*

This concept can, but does not have to, include the one of *customisation*, which concerns the modification of the technology itself. Customisable systems can be modified and adapted to be used in different contexts and often in a collaborative way, whereas appropriation, for Dourish, consists in the adoption of the practice itself. Appropriation may certainly involve customisation, but may also simply mean using a technology for purposes other than the one thought by the designer. By analysing the features of Placeless, an information sharing system, Dourish proposes a guideline to achieve appropriation, in which he focuses on the importance in allowing multiple perspectives of the same information and on the visibility of the correlation between actions and their consequences in the system.

Carroll [Carroll, 2004] looks at the appropriation of a technology as a part of the design process itself. The **Technology as Designed** by programmers is completed by users as they *take possession* and appropriate it for their activities over time. Like Dourish, she defines appropriation as a more fundamental process than simply modifying and configuring a system: it is a *mutual adaption* in which reshaping the features of the technology by the users for unanticipated purposes corresponds to a reshaping of the practice *by* the technology.

On the same wavelength, Alan Dix, in his "Design for Appropriation" [Dix, 2007], proposes both a list of advantages of systems which encourage appropriations as well as guidelines to achieve so.

Properties Advantages of technologies that support appropriation have been summarised by Dix into three major properties:

- **Situatedness** The property of a technology to change the environment in which it is situated. The designer cannot expect to know the environment in which a certain technology will be used.
- **Dynamics** As the needs of environments and users change over time, if the technology is designed for a specific work group and environment then it is more likely that it will become obsolete when such work group and environment change. *Design for use must be design for change.*
- **Ownership** When users appropriate a technology, they develop a sense of ownership. This may be both in using such technology in their own, personal way or by explicitly modifying it to reflect their taste and needs.

The technology appropriation cycle Appropriation is a process which happens over time. Based on the complexity of a technology and the amount of time users spend using it, this process can vary from few days to several months or even years. Caroll analyses this process by subdividing it into three distinct levels. On the *first level*, interaction is entirely based on the features offered by the technology and driven by the expectations that users have on such technology, created by both how it has been marketed and by the tacit knowledge they have about it.

During the *second level*, the technology is then fully explored and adapted to afford and constrain users' activities. It is on this stage that the malleability of the technology will be evaluated to test whether it can fit practical needs. Users will either reject or appropriate the technology consequentially.

Finally, *level 3* will be achieved over time when the technology will be stabilised so to become itself part of users activities. This is what Carroll defines as appropriation and its result as **Technology in Use**.

The guidelines in HCI Gasson [Gasson, 2003] reflects on the unawereness of users' *tacit knowledge embodied in current technologies and practices*, arguing how *technologies are designed around a set of assumptions concerning what work process are required and how they will take place that are often simply wrong.* Along the same lines, Dourish observes that users, and not designers, establish coupling with a technology, whereas Carroll claims a need for **malleable technologies** so for a Technology in Use to reflect users needs.

Coherently with this thread of thoughts, Dix develops a series of guidelines for designers in order for their technologies to allow for users appropriation:

- **Allow different interpretations** and avoid that everything in the system has a fixed meaning.
- **Provide visibility**. Similarly to Dourish, Dix stresses the importance on the relationship between users' action and system reaction, which has to be easy to understand and remember.
- **Expose the intentions** of the system instead of hide them, so to encourage users to leverage on them in the appropriation process.
- **Support not control**. Instead of designing the system to accomplish the task, the system should be designed so that the task *can be done*. This is another point in common to what Dourish, in his Where The Action Is, describes as *an informal assemblage of steps rather than rote procedure driven by the system.*
- **Plugability and configuration** The system should allow to plug-and-play different components of the system for reconfiguration.
- **Encourage sharing** of the modifications made to the system. This point is closely related to the sense of ownership that users develop when they appropriate a technology.

The topic of appropriation in Digital Music Instruments The topic of appropriation has been transposed to the domain of DMI design as well. Following the already mentioned Jorda's "micro-diversity" to describe the variations that could occur within a given piece while keeping it recognisable, [Gurevich et al., 2012] Gurevich et al. explore the use of constraints to encourage novel forms of appropriation. By running a user-study where participants had to prepare a short performance using a highly constrained music instrument, one with only one single momentary pushbutton that generates a single tone, authors analyse how the development of a style in playing a musical instrument depends not only on performers' skill, but also on constraints that can be physical, conventional and imposed by the designer.

These results are successively confirmed by Zappi and McPherson [Zappi and McPherson, 2014], where two different versions of a highly constrained music instrument (a cube containing a touch/force sensor, an embedded computer and a speaker) were proposed to 10 musicians: the first version allowing for a one single degree of freedom interaction and the second allowing for two. The study confirmed that the more constrained the instrument is, the more musicians are encouraged to develop personal techniques and styles, leveraging on its *hidden affordances* strongly affecting the appropriation process. Overall, the paper establishes the formal transposition of the concept of appropriation into the world of Digital Music Instruments.

So, should music interaction be easy? Transposing HCI concepts to the world of DMI is not straight forward. Using a software to achieve a specific task and playing a musical instrument are two very different matters and the application of these HCI guidelines in the world of interaction with music technology represents a real challenge.

Wessel and Wright [Wessel and Wright, 2002] argue that if an instrument has been designed purely to be 'easy to play', it risks to be seen as a toy without challenging users to engage for long-term practice to develop performance skills. Jordà [Jordà, 2004b] formalises these insights introducing concepts such as 'efficiency', 'apprenticeship' and 'learning curve'. In *'Should Music Interaction Be Easy?'* [McDermott et al., 2013], McDermott et al. observe that musicians, when learning to play a novel instrument for the love for music, are in a very different mind-set from the one of software users, impatient to accomplish their tasks as soon as possible. The work stresses on the importance of *difficulties* in the design of a musical instrument. Difficulties permit to differentiate an amateur from a skilled performer, allowing for virtuosity and for communicating efforts and emotions to the audience. The authors therefore distinguish these performance challenges from the 'peripheral and technical tasks', which should be as easy as possible. For example, whereas performing a guitar solo is a task which should not be simplified, the act of tuning the strings of the guitar should be as immediate as possible. Similarly for digital instruments, it is a good practice to provide an extensive set of presets to shortcut the time required for musicians to get close to the sounds they are looking for, offering a good starting point to more advanced customisations.

McDermott et al. forge the term *layered affordance* to describe the feature of a system to offer different degrees of difficulty based on users' skills and experience. This delicate equilibrium between providing an initial reward when performers approach an instrument for the first time and the possibility of reaching a complex and unique interaction with the instrument with practice is perfectly summarised in the already cited work of [Wessel and Wright, 2002]: *low entry fee with no ceiling to virtuosity.*

This thesis supports McDermott's considerations, arguing that adaptable DMIs should be compatible with these desiderata. The complex act of modifying an instrument should be seen, adopting McDermott's terminology, as an easy, immediate and possibly pleasant task, which should ideally not distract the players from their central activity of creating music and embodying a transparent interaction with the instrument.

2.1.4 Conclusions

This section discussed the principles of embodied interaction and highlighted its relevance in the world of music instruments. This interaction happens when a reliable coupling is established and players can focus on the aim of their actions (the creation of sound) and not on the interaction with the instrument itself. I then highlighted the interest of customisation for acoustic and digital music instruments, pointing out its technical complexities and challenges in integrating it in an embodied interaction workflow. Although customisation has been shown to be a relatively common practice in both acoustic and digital music instruments, little attention has been put on targeting it as a design goal and a feature of the final instrument itself.

Although extremely promising, there is a risk in the usability of malleable technologies as their flexibility can bring to a lack of constraints and absence of a guidance on the use of the technology. As observed by Carroll in [Carroll, 2004], a major challenge of designing for appropriation is "to construct a technology that is malleable but still embodies and represents a theory of use that is accessible to users".

It is the goal of this work to study techniques to provide such a guidance and a reward to users during the appropriation process. Bearing this in mind, we move now to the next section which reviews different techniques to help implementing these concepts in music technology.

2.2 Gestural interfaces

2.2.1 Gestural interfaces and gesture classification

Gestural interface is a generic term which refers to human-computer interfaces in which users can interact through their movements, either using the whole body or using only a part of it such as hands. By leveraging on our natural skills in controlling our body, gestural interfaces aim to improve and simplify the interaction with digital systems. In order to avoid confusion with the term *gesture* [Jensenius, 2014], this thesis uses this term to refer to the way for a performer to use body motion in order to control the computer expressively and conveying meaning of interaction [Jensenius et al., 2010]. For example, turning the knob of a MIDI controller involves the usage of an hand to control the computer, but does not convey the meaning of the interaction through the movement itself: the knob could be associated to any parameter and the meaning of the action is not contained in the movement itself (a *weak* relationship following Norman's terminology [Norman, 1998]).

Gurevich et al. [Gurevich and Fyans, 2011] distinguish between instrumental and non-instrumental relation between the performer and the instrument. Authors compared audience's perception whilst assisting to a performance based on a theremin and one based on an unknown DMI in which the gestures of the performer had no clear association with the sound generated. For Gurevich et al., the best way to appreciate a performance is to have an understanding of the functioning of the instrument in order to comprehend the structure of the piece and, as a consequence, the stylistic choices of the performer. Interestingly, they observe that when the audience is not able to associate a clear meaning to the gestures of the performer, they start to focus on other aspects of the piece such as the sound itself, and attempt to build the structure mentally during the fruition of the piece. Performers gestures build the essence of a performance experience, enhancing diversity and uniqueness from a performance to another and motivating audience to watch live concerts in comparison to listening to music at home.

Gesture classification is a particular task of gestural interface software which consists in analysing gesture input from the users, identify specific gesture templates and coupling them with a given meaning. This section considers some gesture classification technique which provides interesting features with respect of the guidelines reviewed in the previous section. Section 2.2.2 discusses an approach to the design of gestural interfaces called *Interactive Machine Learning* and how this relates to the goals of our work. Section 2.2.3 reviews existent works and techniques which relate to a particular form of gesture classification called *continuous classification*. These approaches offer the technical basis for the system presented in chapter 3.

2.2.2 Interactive machine learning

The classification systems that are used in most gestural interfaces are often designed to recognise specific gestures that are hard-coded in the system by software programmers and they cannot be changed by the users. Users are therefore asked to learn these gestures in order to be able to use such systems.

In this work, however, we are interested not only in classifying different users input, but also to give to users the possibility of personalising their systems in order to meet their needs and flavours. The ability for users to define their own gestures has been demonstrated to be important in previous work [Wobbrock et al., 2009]. The field which studies techniques that allow and guide users in this learning process was firstly identified by Fails et al. as Interactive Machine Learning (IML) [Fails and Olsen, 2003b].

Different systems aim to accomplish IML in their design. Crayons [Fails and Olsen, 2003a] is a system for computer-vision classification of images that explicitly encourages users to iterate through the design process by providing immediate feedback on system performance based on the training set. Exemplar [Hartmann et al., 2007] is a tool for rapid prototyping different associations between sensor input and application logic by demonstration. It proposes techniques to both manipulate the input directly and through pattern recognition techniques to enable designers to control how users' examples are generalised to interaction rules. Ruiz et al. [Ruiz et al., 2011b] presented the results of a guessability study where they asked to participants to define motion sensors-based gestures using smartphones. This work demonstrates how users share similar ideas in mapping between motion gestures and specific associated meaning. This information was then used to introduce a gesture vocabulary and to describe a motion gesture set inspired by such results. Lü et al. [Lü and Li, 2012] presented a system for multi-touch screens that allows application developers to program gestures by providing few examples, showing that the system lowers the threshold of programming multi-touch gestures. Magic [Ruiz et al., 2011a] is an accelerometers-based gesture designer tool that graphically plots recorded gestures and makes available video of the designer while performing them. It also gives feedback about the quality of the training set by testing it against a corpus of everyday activity. The Wekinator [Fiebrink et al., 2011] is a software package that aims to make the Weka library more accessible for non-experts allowing users to develop realtime applications, particularly in the music domain. It provides a graphical interface to help users in selecting and configuring different algorithms, as well as allowing users to train, classify, view and correct the classifications.

These systems provide different and interesting approaches to the task of Interactive Machine Learning. However, most of them are discrete recognition systems and consider gestures as a whole indivisible entity. Recently, Caramiaux et al. [Caramiaux et al., 2014a] presented a study that offers a user-defined mapping between the gesture and the sound centred on the listening (or *evocation*) activity. Users listen to a sound first, and then mime it gesturally in order to train machine learning algorithms and therefore defining the mapping.

The next section reviews works in which continuous classification is seen not only as a feature which allows for a richer and more expressive control of the target application, but also to provide a continuous and more subtle feedback to users. This will prepare the ground for the work presented in chapter 3, where it will be used to guide users through the customisation process of the system, making it easier and more enjoyable.

2.2.3 Continuous gesture classification

Digital gesture interfaces are systems composed by a set of hardware sensors and a software program, which extracts useful information from sensors data to control the target application. Sometimes, this information is *unclassified*. For example, a depth-camera based video game may use the position of users hands to control the position of the hands of the protagonist of the game to catch objects on the screen.

In other cases, instead, these information are *classified*: sensors data are used to establish whether or not users performed a specific gesture, which can be associated with a predefined label and sometimes a semantic meaning. Referring to the same example of the depth-camera based video game, classification systems may recognise when users perform a circle with their hands and trigger a specific action any time this happens, such as for example pausing the game. We can refer to systems that adopt this approach as *classification systems*.

Usually, classification systems treat these gestures as whole, indivisible entities, and their implementation is designed to control discrete events once a given gesture is completed. Several systems that allow for real-time recognition have been proposed (for recent reviews see ([Turaga et al., 2008] and [Mitra and Acharya, 2007]). Whilst many systems operate in "real-time", their output remains essentially discrete quantities, i.e. the gesture labels. Wilson and Bobick proposed to extend the recognition task with parameters describing gesture variations for the creation of adaptive systems ([Wilson and Bobick, 1999]. We report here more specifically systems that were designed to provide users with a continuous flow of information characterising their input gestures.

Visell et al. [Visell and Cooperstock, 2007] described a system based on particle filtering that tracks multiple hypotheses about user's input, and can display predictions of future trajectories. This system, targetting applications in physical and neuro-rehabilitation, was designed to allow for a close-loop between the action and feedback given to the user. Williamson [Williamson, 2006] outlined a system for displaying information regarding uncertainty in the continuous recognition task, provided by Monte Carlo sampling methods, and its application for controlling granular synthesis as auditory display. Rodriguez et al. [Portillo-Rodrguez et al., 2008] presented a camera-based system based on Probabilistic Neural Networks and Finite State Machines that allows for the comparison in realtime of Tai-Chi movements between a student performance and that of prerecorded ones by a teacher. The systems generates spatial sound, vibrotactile and visual feedback based on the difference between the student and teacher gestures.

These systems provide continuous, real-time classification of users input and the information they provide can therefore be used to allow users for continuous and synchronous control of a target application. However, the systems listed above require quite complex training in order to be used, as they need to acquire a large dataset of gesture examples beforehand. Therefore, datasets are usually recorded from software developers beforehand and cannot be customised by end-users.

Bevilacqua, et al. [Bevilacqua et al., 2007, 2010] developed a system called Gesture Follower that is designed to continuously output information about the gesture speed and similarity measures relative to a set prerecorded exemplars. One of the features of this system is that it requires only one example per gesture and it is therefore easier for end-users to personalise the system by recording their own ones. This system has been used in artistic contexts for music and dance [Bevilacqua et al., 2012], and in particularly in music and dance pedagogy. This approach will be discussed in details and it will be combined with the techniques described in next section to implement the application described in chapter 3. Recently, Caramiaux et al. [Caramiaux et al., 2014b] presented an extension of the Gesture Follower called *Gesture Variation Follower* (GVF) that, whilst following the gesture, is capable of tracking its temporal and geometric modifications and adjust its parameters incrementally so as to adapt to variations of the gesture such as speed, amplitude and orientation.

2.2.4 Conclusions

This section reviewed some machine learning tools which are used for the implementation of gesture interfaces and are relevant for this work.

Subsection 2.2.2 introduces the field of *Interactive Machine Learning* and reviewed related systems and techniques. The systems reviewed had the common goal of encouraging gestural interfaces users to define their on gesture vocabulary in clear and effective ways. Systems described in this section provide information about users action only once they have already been performed, limiting in this way the feedback that the system can provide to users during the performance itself. However, as saw in subsection 2.2.3, different techniques exist to perform classification of users action continuously over time rather than only once the gesture has been completed.

As we will see in chapter 3, this information can be used to design instruments which provide continuous feedback to users regarding how their actions are interpreted by the system while such actions are performed, encouraging users to interact with the system and particularly allowing for *visibility* and *expose of intentions*, which as seen in 2.1.3 are important features for customisable systems.

2.3 Tangible Acoustic Interfaces

Section 2.1 introduced the concepts of *embodiment* and *reliable coupling*, analysing their importance in the interaction with Digital Music Instruments so as to make interaction with technology more spontaneous and effective. In his book Embodied Music Cognition and Mediation Technology Leman [2008], Marc Leman stresses the difference between music as represented by technology (encoded physical energy) and by humans, which involves personal experiences, interpretations and significations. The author proposes that technology can be employed to extend the capabilities of human perception to mediate between the musical experience and the sound itself, exploring its possible applications both in music instruments interaction and music information retrieval. Dix's concepts of 'visibility' and 'expose of intention' find their correspondence here, where the concepts of *transparency* is defined: *Transparent technology should (...) give a feeling of non-mediation, a feeling that the mediation technology disappears when it is used.*

In the last decade an entire family of user interfaces was born, motivated with the idea of leveraging users pre-existing knowledge about the physical world around us and augment it through digital technology. They take the name of Tangible Acoustic Interfaces (TAI) and are a branch of Tangible User Interfaces (TUI) which considers vibrations that are provoked by touching a solid object as the input of the system.

The term TAI has been employed for the first time in the European project IST-507882 TAI-CHI (Tangible Acoustic Interfaces for Computer Human Interaction) [Crevoisier and Polotti, 2005] [Polotti et al., 2005b], with a clear interesting in applications to physical everyday objects [Crevoisier and Bornand, 2008b].

The goal of this section is first to highlight the general properties of TUI and tangible interaction 2.3.1. We then review some existing tangible music applications, both generic (TUI) and with audio input (TAI), preparing the terrain for the work presented in section 4. Finally we deep into the techniques behind TAI and analyse their features against the goals of this thesis.

2.3.1 Tangible interaction

As observed by [Jacob et al., 2008], the power behind TUI relies in the ability to leverage on users pre-existing knowledge about the physical world, their bodies, the environment around them and the way they communicate with other people. In the field of our work, the design of Digital Music Instruments, Essl and O'Modhrain [Essl and O'Modhrain, 2006] observe how people are used to associate given interactions with the physical world with specific sounds. By preserving the familiar tactile aspect of such interaction, authors claim, allow performers to take advantage of the *tacit knowledge* they have regarding the properties of the world around us.

Tangible User Interfaces draw their strengths by representing digital information with tangible and direct artefacts, serving users with *parallel feedback loops* [Shaer, 2009] in terms of input and output paradigms: haptic feedback serves users that a certain action is completed, as well as providing physical constraints and guidance, whereas digital, often multimodal feedback informs users on when and how their actions has been processed and interpreted by the system. Furthermore, such actions are often not limited on a two-dimensional space as it happens with classic graphic users interfaces and interaction can become three-dimensional and two-handed, as the one with the physical world around us.

Perhaps one of the first formalisation of the design of new tangible user interfaces as a distinct research field has been presented in 2001 by Ullmer and Ishii in [Ullmer and Ishii, 2000]. In this work, authors compare classic Model-View-Control (MVC) model of GUI-based interaction with a new one that they call Model-Control-Representation (MCR),in which they propose to eliminate the distinction between input and output devices.

Paul Dourish, in his book "Where the Action Is", carefully reviews properties of tangible interaction and, also by analysing commonalities with the field of Social Computing, stands the ground for the definition of both embodied interaction and appropriation reviewed in 2.1.1. For Dourish, by finding coherent metaphors for tangible data representation and therefore by studying relationships between actions and meaning, tangible interaction provide promising tools for embodied interaction.

2.3.2 Musical Tangible User Interfaces

I review here the branch of TUIs which has proved to be one of the most successful and, moreover, the most relevant for this work: music applications. The explanations of the success of music TUIs over the last years are manifold. Jordà [Jordà et al., 2007] claims TUIs offer several features which are important for music interfaces: *the support of collaboration and sharing of control; continuous and real-time interaction of multidimensional data; and support of complex, skilled, expressive, and explorative interaction.*

[Shaer, 2009] groups music TUIs into four high-level approaches: fully controllable sound generators or synthesizers, sequencer TUIs that mix and play audio samples, sound toys with limited user control, and controllers that remotely control an arbitrary synthesiser. Systems based on audio as input signal will instead be discussed in section 2.3.3.

The most common model for music TUIs is the tabletop. Tabletops TUIs are table-like interfaces in which a given set of objects on a table are tracked using computer vision techniques and visual feedback is provided by projecting images onto the same table. The first table-like musical TUI that has been published is the music installation Smallfish [Fujihata et al., 2000], where a visual score environment, which generates the music, could be manipulated by moving and tilting plastic rectangles on the table. Audiopad [Patten, 2002] is another table-like TUIs in which users can perform several operations on a set of samples and drum loops by placing tangible tokens onto an augmented surface. New samples can be dragged onto the surface from a menu on the rim.

The most famous music table-like TUI is without any doubt the Reactable [Jordà et al., 2007], with has been seen in action by millions of people on YouTube and even in rock stadiums after being used as part of the set of artist Björk for her 2007 world tour. The Reactable system allows to visual program and dynamic patching moving and tilting tagged objects on the table. Every object has a specific function and connections between two compatible objects are suggested as their input and output slots automatically attract each other through proximity. As Jordà claimed, the *foremost goal designing the Reactable was to design an attractive, intuitive and non-intimidating musical instrument for multi-user electronic music performance* that is engaging from the first minute but also is complex, subtle and allows for endless variation. We can spot here a clear analogy with the *"low-entry-fee vs no ceiling for virtuosity"* desideratum proposed by Wessel and Wright and reviewed in section 2.1.2.

There are few examples in which TUIs are embedded into objects brought from our daily life in order to build digital music instruments. Music Bottles [Ishii et al., 2001] is an installation which uses glass bottles as a controller to turn on and off different music tracks of different music styles using small electromagnetic resonator tags placed around the opening of each bottle [Paradiso and Hsiao, 1999]. The Squeezables [Weinberg and Gan, 2001] is a music TUI that allows a group of players to perform and improvise musical compositions by squeezing and pulling six get balls mounted on a small podium. The audio shaker [Hauenstein, 2004] is a container which allows to capture, shake up and pour out sounds. Sounds are recorded by removing the cap, pronounce something near the container and then close it again. Then, shaking and tilting the object will provoke different audio effects on the playback of the sample. Cubed [Stanley] is a music step sequencer controlled by a physical Rubik's Cube. By manipulating the colours on the cube, users generate different sound algorithms within the sequencer. The Mixmaster [Niinimaki, 2009] is a vintage household mixer with a wireless sensor and an Arduino circuit board embedded. By moving the mixer horizontally, users can scratches pre-recorded samples while a button is used as a turntable crossfader. The 'MO objects' [Rasamimanana et al., 2011; Schnell et al., 2011] are a set of tangible objects and software modules which are designed for musical interaction and performance. As their design and properties are particularly relevant for our work, the MO objects are discussed separately in section 2.3.3.

The MaKey MaKey [Collective and Shaw, 2012] is a device that allows users to easily employ everyday physical objects as computer keyboards. A set of high resistance switching called 'Alligator' detect when users touch an object and triggers a specific letter of the keyboard. The MaKeyMakey is highly generic as it allows to control any kind of musical and non-musical application whilst being extremely simple and easy to use. As it doesn't allow for continuous control, it is particularly suited for applications that requires discrete control events. LittleBits Bdeir and Ullrich [2011], an opensource kit of pre-assembled electronics that snap together with tiny magnets to easily connect sensors and software triggers, and PatchBlocks [Pat], an hardware audio effect that can be easily reprogrammed through a bespoken visual language, are other examples of a rapid emergence of kits to easily build custom tangible interfaces.

TUI offers a very promising approach as they are generally more engaging that

GUI-based systems and less difficult to learn. They take advantage of our natural skills in manipulating physical objects and therefore they are less intimidating for non expert users. They are also particularly suitable for scenarios in which shared control is aimed, as this can easily be distributed through the interactive area. Control bandwidth can potentially be high as it is possible to use two hands per person and most of music TUIs allow for continuous control and feedback.

These systems are often presented as ad-hoc finalised combinations of hardware and software, often adding physical constraints to guide users in performing series of actions pre-designed by programmers. Although the immediate effect of this type of design is to make systems straight forward to use in the very immediate, this gives small room for customisation and adaptation of such systems for individual practice. The next section focuses in a particular branch of TUI which provide interesting features for this regard: Tangible Acoustic Interfaces.

2.3.3 Musical Tangible Acoustic Interfaces

There are several tangible interfaces that use audio as an input signal to create digital music instruments. This section groups them into two main categories: the systems that use physical objects as pure controllers and are linked to independent sound engines and the ones in which such the sound output strongly depends on the nature of the particular physical object employed.

Physical objects as controllers The idea of using audio as control signal to trigger music events is far from being new. Electronic drum kits that can detect onset in the signal and trigger single MIDI event correspondently have been commercially available for decades. The limit of their expressivity however is fairly evident and have been perfectly summarised by Miller Puckette [Puckette, 2011]:

> *(Discrete triggers) are far less expressive than instruments that transmit or process the vibrations themselves; for instance, sliding a brush over a drum trigger isn't likely to produce anything useful, whereas doing the same thing on an instrument that operates directly on the audio signal from the contact microphone has the possibility to create a wide range of useful musical sounds.*

In more recent years various systems started to use audio as control signal in order to adapt the sound synthesis continuously over time, obtaining a much wider nuance of sounds and expressivity compared to MIDI triggers.

A very interesting example of a musical TAI is given by Stane [Williamson, 2008], a small palm-shaped device with an embedded piezo-transducer and a multitude of engineered textures on its surface. The system allows for *continuous classification* of user touch among the following four classes: Scratching circular front clockwise, Scratching dimples on right side, Scratching tip with fingernail and a Miscellaneous noise class. This is done through the following technique. The system analyses the vibrations that are sensed by rubbing different areas of the surface and computes an FFT of a windowed signal composed by 512 samples and overlap of 7/8. The spectrum is then rebinned so that bins are four times their original size. Each one of the such four classes is trained on two minutes of input data which is recorded as both motion, pressure and grip postures. The feature vectors so obtained are then classified using multi-layer perceptron technique. Stane is described by authors as capable of reaching 75% accuracy at this stage. However, these results are then passed to a dynamic system which smooths out the fluctuations in the classifier to improve better results. Authors provide a use-case application in which they use Stane to control different functions of a music player.

A second system which we can classify under this category is the Table Recorder [Gmeiner, 2007], a sonically augmented table in which any time the table is touched by users bare hands or through an object, one of the actuators embedded under the table is triggered to produce a different sound and through different modalities.

The PebbleBox [O'Modhrain and Essl, 2004] already mentioned in 2.1.2 is another system which falls in this category. The physical object consists in a box of grains made by several, arbitrary materials. The system analysis the audio generated by physically touching and moving such grains in order to control a granular synthesiser. The main idea behind this system is to build a direct link between the haptic feedback of touching physical grains and the "granular" nature of the associated sound engine.

Similarly, in Tactophonics [Cook and Pullin, 2007] a granular synthesiser is controlled by spectral parameters generated from a piezo-transducer that users can attach to any object. In both PebbleBox and Tactophonics no actual classification is performed and audio input features are directly mapped onto sound synthesis parameters continuously overtime. [Cook and Pullin, 2007] also provide an interesting design research in musical affordance by using sounds as control signals.

The 'MO objects' [Rasamimanana et al., 2011; Schnell et al., 2011] are a set of tangible objects and software modules designed for musical interaction and performance. Several wireless blocks embed accelerometers, gyroscopes and contact microphones and send sensor values to a computer that hosts several MaxMSP [cyc] music applications. Such blocks are designed to be easily combined together and embedded within everyday objects to transform them into musical games. MO moves a step forward in the direction of this work, as they are extremely customisable and users can embedded them into objects of their choice and use them as controllers for different sound games presented as MaxMSP patches. See [Rasamimanana et al., 2011] for examples of a sonically augmented ball and chess games.

All the systems described here combine tangible, haptic feedback with continuous control allowing physical objects to become expressive controllers for ad-hoc sound engines.

The shared goal amongst these systems is the one of using the subtile nuances of the audio signal coming from piezos in order to enhance the control bandwidth and expressivity of the instruments.

Controllers and sound source as unified entities In all the systems mentioned so far, the objects employed as TAI act as controllers able to offer a very intuitive approach for complex sound processing and music games. However, as Delle Monache et al. noticed [Monache et al., 2010], "the produced music still remains detached from a real source: the sound controlled or generated via the interface manipulation is still not the sound of a physical object".

Steve Mann, one of the founder of the concept of Natural User Interfaces, in 2007 introduced the concept of physiphones [Mann, 2007], DMI in which both the user interface and the sound production medium are based at least in part on natural physical phenomena. Other instruments by the same author such as the poseidophone and the hydraulophone use natural material such as water or gas both as interface and as basis for the sound output. In 2012 Mann extended this concept by introducing physiphones in which the final sound delivery is also physical, i.e. it is propagated through the instrument itself instead as through headphones or loudspeakers.

In 2005 Crevoisier et al. [Crevoisier and Polotti, 2005] proposed several TAI implemented by combining together a series of techniques: TDOA, time reversal, Active Acoustic Holography, amplitude follower, pattern recognition and spectrum analysis. As these techniques aim to retrieve different information about users interaction, they are complementary and thus can suit well together. However, as previously mentioned in section 2.3.4, some of them require complex training for the specific object that has to be used and therefore this process cannot be easily undertaken by end-users.

The Sound of Touch [Merrill and Raffle, 2007; Merrill et al., 2008] is a sound installation developed at MIT in which a physical hand-held wand is augmented with a piezo-transducer and a button. The button allows to trigger the recording of an audio sample, for example a word or a sentence spoken by the user. Then, the user can brush and scrape the wand against a range of different material on a surface. The sound produced by the wand, and picked up by the transducer, is digitally convolved with the pre-recorded sound using a particular technique proposed by Aimi et al. [Aimi, 2007]

On the same wavelength, Schwarz et al. [Schwarz et al., 2014] presented a system which combines piezo live-input and concatenative sound synthesis. Through the CataRT system [Schwarz, 2007], a database of prerecorded audio material is divided into small segments, pre-analysed and displayed in the 2D surface of a touchscreen. The player can select the grains in realtime to trigger their playback and the resulting sound is convoluted with the incoming signal from the piezo transducer.

Audio physical modelling, the ensemble of sound synthesis technique that aims to simulate the sound or real-world objects, is a large subject that touches various academic fields including psyco-acoustic, music acoustic, sonic interaction design and DMI (see [Rocchesso and Fontana, 2003] for one of the most famous references on the argument). Audio physical modelling has been employed in various TAI has well. For example Delle Monache et al. presented the Gamelunch [Monache et al., 2010], a work in which force sensors, accelerometers and piezo-transducers have been employed to sonically augment several kitchen objects using physical model [Adrien, 1991] and waveguide [Smith, 1992] synthesis. Authors propose seven pre-designed sound presets that can be associated with different objects (in their application example these are: a fork, a knife, shakers, a decanter, a sangria bowl, a salad bowl and a tray)[Polotti and Monache, 2008].

Puckette [Puckette, 2011] presented an audio-driven digital resonators model, used to augment a ceramic tile attached to a piezo-transducer to create a percussive instrument, using the incoming audio as exciter for a nonlinear reverberator. Oppositely to Mogees, the system described in chapter 4, here the signal is deconvolved from the impulse response of the physical object, so the resulting sound is as independent from it as possible.

Other notable instruments that combines audio input and physical modelling are the Kalichord [Schlessinger and Smith, 2009], which feed a physical string model with plucks from acoustic tines; Edgar Berdahl and Julius Smith's tangible virtual vibrating string [Berdahl and Smith, 2008], a guitar-like instrument where multi-axis pickups are used to excite a two-axis digital waveguide virtual string; the Chameleon Guitar [Zoran and Paradiso, 2011] is a guitar with an embedded DSP for audio-driven digital resonators that allows users to easily replace the physical objects used as exciters; the SpectraSurface [Hattwick et al., 2014] is a recent percussive instrument which maps various audio spectral descriptors to the sound synthesis; and finally the Korg Wavedrum is a commercially available interface that uses contact microphones attached to the skin of a snare to send audio signal to a synthesis engine.

2.3.4 Technologies

Using acoustic vibrations as input paradigm for tangible interfaces offers several advantages. First of all, most of the objects we use in our daily life emit vibrations and therefore they are suitable to became TAIs, which offers interesting premises in relation with the Situatedness desideratum introduced in section 2.1.3. Piezoelectric transducers, the sensors used to pick up vibrations, cost few pennies, are generally tiny (in the order of few millimetres) and non intrusive and can easily be embedded or placed on top of objects without altering their structure. The emitted electric signal can be sampled at audio rate and its frequency range is contained in the audible domain. Thus, it can be received by any existing commercial sound card, laptop and mobile phone.

Different techniques take advantage of wave propagation in solids in order to extract useful information about users touch. This section reviews the main techniques used to implement TAIs, discussing both supervised and unsupervised approaches designed to achieve different results.

Time difference of arrival The first technique reviewed in this section is called Time difference of arrival (TDOA) and it uses an array of piezo-transducers conveniently arranged onto the surface as sensors for the system. This technique is based on a two-step procedure. First, the system has to be trained by each couple of sensors and time delay estimation has to be recorded. Such information is then processed to build hyperbolic curves that calculates, for each couple of transducers, the estimations of the locations that correspond to the recorded delays. When this process is completed, such curves are intersected one with the other in order to locate the position of new inputs. [Rindorf, 1981] [Haykin, 1985] .

Although this method has been designed for flat surfaces, the same principle can be used to retrieve information over non-flat surfaces. As explained by Polotti et al. [Polotti et al., 2005a], one of the most crucial problem in this scenario is given by the dispersion of the phase velocity which occurs with wave propagation in solids.

According to [Shaer, 2009], the first system to use an array of sensors to implement a tangible user interface was the PingPongPlus developed at MIT in 1999 [Ishii et al., 1999] [Checka, 2001]. PingPongPlus consists in 8 piezoelectric sensors installed under a pingpong table to estimate the position of the hit of the pingpong ball on the table. A video projector, mounted on top of the table, projects visual animations based on such information.

The Responsive Window [Paradiso et al., 2002] [Paradiso and Leo, 2005] , the Know-Activated Browser [Paradiso and Checka, 2002] and the Telephone Story installation [Beltran, 2001] are other MIT projects that use similar techniques combined with video projection to augment a glass panel window into an interactive surface.

The Tangible Acoustic Interfaces for Computer-Human Interaction (TAI-CHI)

European project [tai, 2009-2011] explored different applications that used this technique [Polotti et al., 2005a]. For example the Percussion Tray [Crevoisier and Bornand, 2008a] is an augmented plastic tray with a piezoelectric sensors at each of the four corners. The four audio channels are sent to a computer that calculates the position of the touch and discretise it among 16 different positions in a grid which are associated to the triggering of 16 different drum samples. Other applications developed in the context of the TAI-CHI project are discussed further on.

Sokolovskis and McPherson [Sokolovskis and McPherson, 2014] employs TDOA to locate the position of a hit in a drum snare with an accuracy of 2cm, using optical sensors installed underneath the skin.

Mimio [mim] is a commercial system that also uses an array of piezo to determine the position of touch.

These systems have the advantage of retrieving the position of the touch in a relatively accurate manner (although not comparable with computer vision based approaches) and low latency. However, the setup process is quite delicate and requites time and expertise. Therefore, this technique requires programmers of the system to perform the calibration beforehand for a specific surface.

Time reversal Another technique used in TAI is the Time Reversal, which leverage on the assumption that an acoustic signal can be reconstructed in its original location by recording the received signals and then sending back the time reversed version of these signals through the medium [Fink and Prada, 2001].

As a consequence, the received signal contains information regarding its source location and *the impulse response in a chaotic cavity is unique for a given source location.* [Pham et al., 2007]

Time Reversal technique, sometime referred as Location Template Matching (LTM), therefore allows to retrieve the absolute position of the touch comparing the incoming peaks with pre-recorded templates.[Ing et al., 2001] Like any supervised classification approach, the procedure is divided into two steps:

1. Acquisition of the impulses (training): various impulses are performed at different locations and corresponding responses are recorded. The duration of the response depends on the absorption of the material and on the energy radiation property of the cavity.

2. Classification: the absolute position of the touch is computed by calculating the cross-correlation between the incoming input and all the pre-recorded templates.

See [Pham et al., 2007] for details about the cross-correlation distance measure.

One major difference of this approach compared to TDOA is that here usually only one sensor is sufficient for flat surfaces, as long as it can be fixed away from a symmetry axis. However with non-flat objects the symmetry axis are not uniques and therefore there can be more identical acoustic responses for different locations. [Crevoisier and Bornand, 2008a] A possible solution to this problem is to place sensors so that they are not on the same symmetry axis. For a comparison between this technique and the previously described TDOA, see [Pham et al., 2007].

Several tangible acoustic interfaces have been built using this approach. For example the Light Globe [Crevoisier and Bornand, 2008a] is a light globe augmented through contact microphones. As the object has a spherical shape, two contact microphones are used for the reason described above. During the training phase, users are asked to tap several times according to a visual feedback provided by a certain number of LED's. When the calibration is done, the system is able to recognise the position of the touch with an acceptable degree of precision.

Time reversal is a powerful technique that allows to retrieve the position of the touch in a wide range of solids. This technique requires a training that is specific for the objects users want to augment. Unfortunately the training process can be quite complex and error prone and needs to be done with appropriate visual feedback. Moreover, the technique allows to describe discrete peaks and it is not directly suitable for continuous control.

Active methods Although not used to implement TAI, it is interesting to mention that techniques based on ultrasound also exist. These techniques, mainly regarded as in-solid acoustic Holography, are called 'active' as they consist in exciting an object with ultrasound and that evaluate the acoustic energy that is absorbed at the points of contact.

As an example, the Cricket System [Priyantha et al., 2000] uses coded ultrasound pulses to locate and identify users in an instrumented room. A listening device is attached to the users hand and calculates its position in the room by measuring its distance to a set of fixed, coded emitters.

[Ciglar, 2010] presents a contact-free instrument that employes an array of ultrasonic transducers arranged on a spherically shaped surface, capable to generate sounds in a single position thus providing vibro-tactile feedback to the performer. By changing the position of the hand, the performer modifies the intensity by which the ultrasonic sounds are delivered back to the interface allowing consequently to retrieve the position of the hand and to use this information to control synthesis parameters.

Hybrid systems Various TUIs combine acoustic input with other techniques such as touch-screens or ad-hoc sensors in order to retrieve information about the touch. For example the multi-touch table [Crevoisier and Bornand, 2008a] is a system that uses two laser emitters placed in two corners at about 1cm above the surface. When users touch the plane, their fingers reflect a light that is detected by an infrared camera placed above the surface.

TapSense [Harrison et al., 2011b] is a touchscreen with a stethoscope installed on the back. While the absolute position of the touch is retrieved with the capacitive sensor of the touchscreen, the touch is also classified using the acoustic vibrations sensed by the piezo. The classification adopts a supervised approach: the model of every class of touch (pad, tip, knucke and nail) is built from examples provided beforehand. Segmentation is done thresholding on the amplitude of the signal, and then an FFT of 4096 samples (43ms at 96khz) is built. Then, several features are used based on the lower 500 bands of the signal, although authors suggest the most expressive range is below 1khz. Classification is then achieved using Support Vector Machine.

Sonically Enhanced Touch [Lopes et al., 2011] adopts a similar approach using normal contact microphones and Bounded-Q analysis as described in [Puckette et al., 1998].

TouchLight [Wilson, 2004] is a system that combines computer vision techniques with audio input (the impact intensity of the surface contact is used to distinguish a tap from a knock).

Scratch Input [Harrison and Hudson, 2008] is another supervised classification system with acoustic input that a modified stethoscope as sensor. However, the main difference of this system compared to the ones previously described is that it has been designed to work with a wide range of surfaces instead of a specific one. The goal of the system is to classify among 6 types of touch without stressing on the idea of retrieving the position of the touch. Therefore, the proposed applications described in [Harrison and Hudson, 2008] have been tested in different walls and surfaces, although performing different trainings for every surface. Authors suggest the technique could be used to achieve continuous control by performing a discrete classification of a gesture to select the task and successively using the amplitude of the incoming signal for continuous control.

Skinput [Harrison et al., 2010] is a recent system that employes an ad-hoc array of 10 wearable bio-acoustic sensors with the aim of using human skin as an interactive surface. These sensors are particularly sensitive to low frequencies (i.e. 25 hz) that authors claim to be an important range to recognise finger tips performed on the skin. Signals are sampled at 5.5 kHz and segmented into individual taps using linear intensity threshold and described through a range of features. Incoming peaks are then analysed and compared with the pre-recorded templates using Support Vector Machine. In order to be used, the system needs to be trained for a specific person by tapping in several locations of his arm.

Temporal patterns In 2012 two TAIs have been proposed that differ from the ones previously described as, instead of segmenting and classifying audio peaks, use their temporality to retrieve unique identifiers. The first one is called Acoustic Barcodes [Harrison et al., 2012]. The system uses a piezo-transducer to sense structured patterns of physical notches that can be printed onto surfaces or objects and, when swiped with a fingernail, produce a train of onsets that can be resolved to a binary ID and thus classified.

The second one has been presented by Ghomi et al. [Ghomi et al., 2012] and uses rhythmic patterns as an input method. The system thresholds the amplitude of the incoming signal and converts it to a binary value (on and off). It measures then the length of every segment dividing it into three quantities: impulse, one beat and two beats. Patterns between 2 and 6 beats long are thus created and classified.

Although this technique sounds very promising to extend the range of users input, it is difficult to imagine how it could be improved to allow continuous control, as the classification can be done only at the end of such temporal pattern.

2.3.5 Conclusions

The systems described in this section demonstrate how Tangible Acoustic Interfaces can be a powerful, fast and cheap tool to retrieve information about the way users interact with solid objects.

I described systems which employ tangible everyday objects as controllers for given sound engines. The main advantage of these systems relies in the fact that, in addition to providing haptic feedback like the other tangible interfaces, they can leverage on users familiarity with everyday objects they have around them. However, as noticed by Delle Monache et al., such systems keep the controller and the sound engine as separate entities and the sound outcome is not directly affected by the particular nature of the object employed.

Existing TAI in which the sound source is the physical object itself have been reviewed. By adopting such approach, these systems propose interesting solution to solve the disconnection between controllers and sound sources present in Digital Music Instruments. Such systems can take full advantage on our knowledge in *sonically* interacting with everyday objects and offer the potential to reach high *visibility* , defined by Alan Dix as the relation between users action and system reaction.

However, the techniques described in subsection 2.3.4 often require complex calibration before they can be used and are therefore embedded in ad-hoc, preconfigured hardware-software systems. Furthermore, most of them focus on discrete classification, i.e. the classification of a temporally segmented input among a fixed number of possibilities, and therefore are not directly suitable to provide continuous control.

This section ends the literature review for this work. Different concepts behind the phenomenon of *customisation* have been discussed, demonstrating its importance in the design of music technology starting from a philosophical perspective and moving them to the world of HCI and DMI. Section 2.2 reviewed some Machine Learning tools which look promising in relation to these aspects. We are now finally ready to apply such tools for the creation of a novel gesture interface for the control of digital media as well as studying its benefits in the interactions with users. Similarly, the issues raised in this last section which relates with Tangible User Interfaces will be addressed in chapter 4 where novel techniques for the creation of a novel music TAI will be proposed.

Chapter 3

Applications of continuous recognition for fluid gesture interaction design

This chapter presents Gesture Interaction DEsigner (GIDE), an innovative application for gesture recognition. GIDE is an interface for an existing gesture recognition algorithm that compares users gestures with prerecorded templates in realtime as they happen. GIDE proposes an interaction workflow supported by an ensemble of visualisation and sonification techniques to make the usage of the algorithm and the calibration of its parameters easier and more understandable for end-users. Rather than learning the pre-defined gestures of others, the resulting application allows users to design their own gestures so making interaction more natural and also allowing the applications to be tailored by users' specific needs. Furthermore, instead of recognising gestures only after they have been entirely completed as happens in classic gesture recognition systems, GIDE exploits the full potential of gestural interaction by tracking gestures continuously and synchronously so allowing users to both control the target application moment-to-moment and also receive immediate and synchronous feedback about system recognition states. By this means, they quickly learn how to interact with the system in order to develop better performances. The chapter describes in details the system and the techniques employed to provide realtime multimodal feedback to the users through a user study with a range of performers and artists.

This chapter is based on: *Bruno Zamborlin, Frederic Bevilacqua, Marco Gillies, and Mark D'inverno. Fluid gesture interaction design: Applications of continuous recognition for the design of modern gestural interfaces. ACM Transactions on Interactive Intelligent Systems (TiiS), 3(4):22:1-22:30, January 2014.*

3.1 Introduction

This chapter presents Gesture Interaction DEsigner (GIDE), an innovative application for gesture recognition. The system can be applied to a wide range of gestural interfaces and aims to enhance their behaviour and encourage users to both obtain a richer control of the system and to customise it for their needs.

The application leverages an algorithm called Gesture Follower by Bevilacqua et al. [Bevilacqua et al., 2007, 2010] for which I worked on for three years. Instead of recognising gestures only after they have been entirely performed as happens in classic gesture recognition algorithms, this algorithm estimates intermediate results of the gesture as it is being performed. Several visualisation and sonification techniques are employed in order to improve the usability and the interaction workflow of this complex algorithm and ease the calibration of its parameters (*low entry fees*) without compromising the accurateness of its results. This ensemble of features allows GIDE to exploit the full potential of gestural interaction for end-users by tracking gestures continuously and synchronously so allowing users to both control the target application moment-to-moment and also receive immediate and synchronous feedback about system recognition states. By this means, they quickly learn how to interact with the system in order to develop better performances. Furthermore, rather than learning the pre-defined gestures of others, GIDE allows users to design their own gestures, allowing the applications to be tailored by users' specific needs.

In section 3.2, the design of the application and the features it aims to provide are presented. The gesture recognition algorithm, the Gesture Follower by Frédéric Bevilacqua et al., is then described explaining the features it provides and the motivations that makes it relevant for the purposes of this work. I implemented the algorithm in C++ and provide in this section a quantitative evaluation of its performances against other more standard approaches using a pre-existing gesture database.

In section 3.3, the interaction workflow proposed by GIDE is explored, alongside with its various feedback techniques. The use of the features described in the previous section are detailed as well as their motivations related to the topics of definition of user-defined gesture vocabularies, parameters tuning and quickly experimentation of the results. Three different DMI based on GIDE for the control and manipulation of digital media are then presented.

Section 3.4 describes in details a user-study with 23 artists from different domains in order to evaluate whether and how the different visualisation and sonification techniques presented enable artists with different backgrounds to understand and successfully use the gesture recognition algorithm employed.

Finally, section 3.5 summarises the main contributions of this chapter as well as reviewing the limitations of the proposed work and outlines future research in this direction.

3.2 GIDE: Gesture Interaction Designer application

This section presents a new gesture recognition system called *Gesture Interaction Designer (GIDE)* (figure 3.1), which allows for recording a series of gestures (the "*gesture vocabulary*"), visualising them and using them as a training set for the recognition of future gestures. GIDE adopts an interactive approach for setting-up the machine learning environment, in order to allow also non-expert users to take advantage of these techniques and deeply customise the application for their own needs.

The gesture recognition algorithm employed is called *Gesture Follower* [Bevilacqua et al., 2007, 2010] and offers two critical results. First, from the moment a performance begins, it enables a continuous estimate to be calculated of which recorded gesture is the one currently being performed. This estimation happens in real-time, moment by moment over time. Second, for each of these potential target recorded gestures, the algorithm provides a continuous estimation of the current temporal position of the performance within each of them. I refer to these features respectively as *real-time gesture recognition* and *real-time gesture following*. This section will also provide a formal evaluation of this algorithm and show how it compares favourably to existing approaches.

Although it has been applied for the control of Digital Music Instruments, GIDE has been designed to be a general purpose application for gesture recognition that can work across different application domains and media. The algorithm makes no assumptions to the nature of the sensors data it receives, and can therefore be used with any stream of data regularly sampled. Thus, this application can be embedded in many digital devices such as video and depth cameras, inertial sensors and so on. Its aim is to enhance the interaction with such interfaces by providing both a higher degree of expressivity while offering the possibility to users to deeply customise the way the system behaves. I have successfully tested GIDE with several devices: motion sensors such as accelerometers and gyroscopes, video camera using image descriptors, sound input (e.g. microphone) using audio descriptors. However, describing in detail all these different applications goes beyond the aims of this chapter and I will instead focus on the use of two distance yet commonplace examples: finger gestures (captured using mouse or tablet) and hand gestures (captured using accelerometer-based motion sensor).

In this section, the desired features of our system are first divided into three interconnected properties, aiming to ease and enhance the process of usage of the system (3.2.2). This establishes the hypothesis of this chapter. In 3.2.3, the results provided by the Gesture Follower algorithm are described and in 3.2.4 they are evaluated against a standard approach called Dynamic Time Warping (DTM), using a 2D gesture database publicly available. Finally, 3.2.5 describes how such features are related to the purposes of this work.

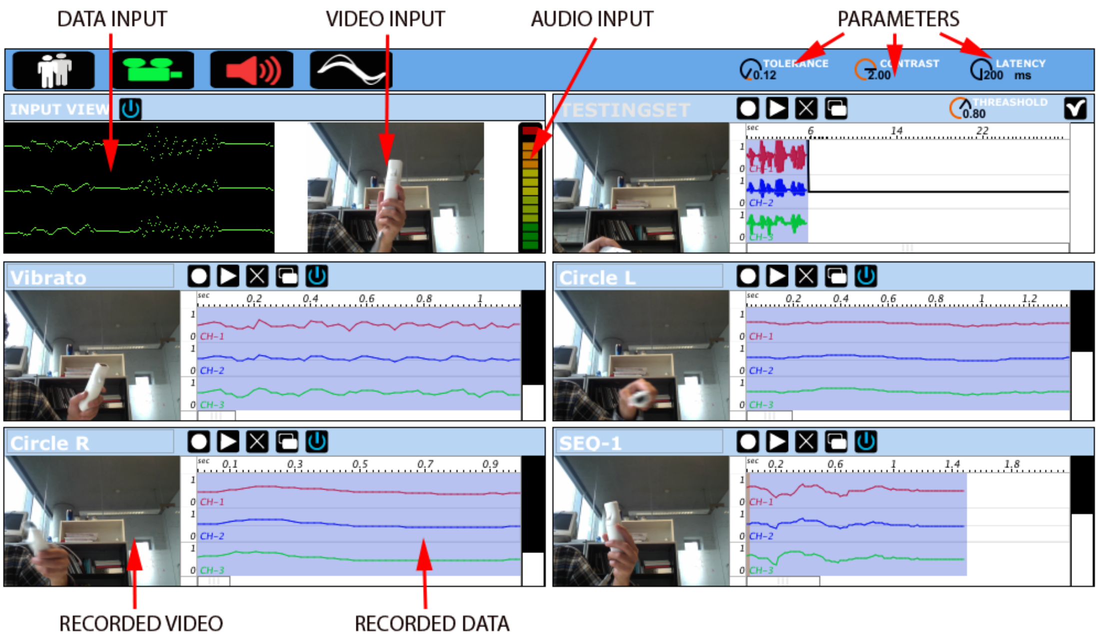

Figure 3.1: Detail of the Gesture Interaction Designer application

3.2.1 Implementation

The current version of GIDE is presented as a stand-alone application for OSX. I implemented the application using MaxMSP [cyc] and the MuBu toolbox [Schnell et al., 2009]. The Gesture Follower algorithm has been firstly presented in [Bevilacqua et al., 2007] as a MaxMSP patch using the MnM toolbox [Bevilacqua et al., 2005]. In 2008, before starting my PhD, I coded the algorithm as a C++ library and successively wrapped it as a MaxMSP object called *gf*. Since then, the object has been used by the IRCAM Real-Time Musical Interactions team and myself for many research experiments and artistic projects, which allowed to improved the code as well as adding new functionalities. *Gf* is employed in GIDE and connected with the various data visualisation and sonification components described in the chapter. Details about the Gesture Follower implementation can be found in appendix A.

3.2.2 Desiderata

I define here the main goals that driven the concept and the design of the GIDE system. For the sake of clarity, they are grouped into four main desiderata, which are sequentially connected one with the other.

Continuous control The general aim of gesture recognition systems is to associate precisely and reliably different user behaviours to semantic meaning. Embodied interfaces should be able to do so continuously, moment by moment through their gestures and synchronously i.e. with a very short and constant latency. I define continuous control here as the property of controlling the target application moment by moment over time *while* users are performing their gestures. This property is in contraposition with *discrete recognition*, implemented by most systems, in which control happens only once the gesture has been completed.

Although for many applications it may be sufficient to trigger discrete events, enabling additional continuous control will extend the range of possible control mechanisms and enhance the peculiarities of embodied interfaces. Expressivity of human body movements cannot be fully represented as a sequence of discrete commands. Continuous synchronisation between user movements and digital processes is necessary to enable expressivity in gestural interfaces. For example, it is generally useful to include the possible modulating effect of continuous changes in gestures occurring between triggering events. This typically allows for the use of important information occurring in preparation gestures, which can be in turn used to anticipate specific control. Using continuous control, it then becomes possible to extend such an approach by extending intermediate recognition results that become available during a gesture performance.

Continuous feedback This quality refers to the ability of interfaces to provide meaningful feedback to users regarding how the system is interpreting their actions. Users need to be able to readily access as much information as possible from the system during any practice or performance episode in order to understand the relationship between their action and the system's response. A certain level of satisfaction or even virtuosity comes when users can perform their actions with sufficient accuracy that they can control the system reliably enough to satisfy their intentions. This feedback should refer ideally to every action of the user with the system, including performing a movement and tuning a parameter of the system to adjust its behaviour. Without this facility it is prohibitively difficult for users to get better at interacting with the system. Moreover, performers need to have the possibility of perceiving this feedback without having to look at the screen so that they are free to focus their attention elsewhere. Finally, and perhaps most critically, users need to access feedback synchronously and continuously over time just as happens when practicing a musical instrument.

User-defined gesture vocabulary This property refers to the ability to let users to define the gestures that have to be recognised by the system.

In most standard gesture recognition systems, such gestures are designed a-priori by the programmer, often to maximise performances, and then hard-coded into the system [Nielsen et al., 2004]. However, the goals of this research is to propose a system which can be easily personalised by end-users and integrated in a specific context. Users are able to define a *personal vocabulary of gestures* specifically for the target application in hand and the environment in which interaction activity will take place. The system provides users with the flexibility to easily modify their gestures as users develop the way in which they want to interact with the system. It enables their user to define gestures which are meaningful and even *metaphorical* to them personally with respect to the response behaviour of that system. If systems do not allow this, then the gestures have *weak* connection with their functions and systems risk being worse than traditional GUIs as users must remember an arbitrary vocabulary of gestures which are not meaningfully grounded in their own individual movements [Norman, 1998]. In such situations, it becomes at least as difficult as remembering an arbitrary set of textual commands and possibly even more difficult if gesture interaction itself is new to the user. I argue that, by providing systems where users are flexibly accommodated so they can link their personal gestures with their intended system response, a *strong* and *specific* connection is created. Indeed this becomes not just desirable but necessary when gesture interaction systems are used as part of an artistic performance such as dance and other contemporary productions. In dance scenarios, for example, gestures must be specifically designed for the choreography of the piece, the specific dancers being involved and even according to the environment of the venue and technology that is available. If the gesture set is pre-defined or limited in any way then it is difficult to see how such systems could be effectively used in performance settings in general.

3.2.3 Background: the Gesture Follower algorithm

I review here the features provided by the Gesture Follower algorithm employed in GIDE. As previously mentioned, a fundamental requirement of this work is to provide feedback about the state of the recognition *during* the performance. Furthermore, users should be able to easily record and edit their gesture vocabulary based on the feedback that they receive from the system. However, most classic machine learning algorithms, such as Hidden Markov Models, need to access to the entire gesture before give a result and are therefore not suitable for our purposes.

Real-time hidden Markov models The Gesture Follower is a modified version of Hidden Markov Models previously described by Bevilacqua et al [Bevilacqua et al., 2010]. Hidden Markov Models (HMM) can recognise sequential data using a probabilistic approach. Series of observations are modelled using a finite number of states, whose transitions are defined by transition probabilities. Each state emits observations based on a probability distribution function. Generally, the HMM's parameters are set through training procedures using a large database statistically representative of all possible variations. However, for interactive gesture design this would require us to collect a large number of users data, with each user repeating gestures many times. This would obviously limit the interactive procedure between designing gestures and receiving feedback on the gestural interface behaviour.

On the contrary, this algorithm makes the learning procedure quick and immediate using an hybrid approach between probabilistic HMM and exemplar based approaches such as Dynamic Time Warping (DTW) that requires only a single gesture example to specify gestures class. The "modified" HMM approach used by the Gesture Follower sets the Markov models from a single example, by associating each example sample data with a state, as shown in Figure 3.2. Such a choice leads us to consider a large number of states and is thus a less efficient decoding computation compared to standard HMM approaches. However, two points should be noted. First, the loss of efficiency was never found limiting in our application. Second, our approach offers the crucial advantage to closely model the data time profiles, which might be lacking in a standard HHM. In particular, similarly to DTW, it is possible to temporally align the incoming data with the original example at the granularity of individual samples. Compared to DTW, our approach allows for real-time decoding during the performance (using the HMM forward procedure) while standard DTW is operated only at the end of the gesture. A similarity measure can be estimated with the same time granularity (at the sample level). These two features together enable the first desideratum: continuous control. Moreover, when performing recognition, the HMM associated with each gesture is evaluated and the one with the highest similarity measure (i.e. highest likelihood) is used to classify the gesture. When performing recognition, the HMM associated with each gesture is evaluated and the one with the highest computed probability is used to classify the gesture.

Learning and decoding As in standard machine learning techniques, the workflow is divided into two phases, learning and decoding. During the learning phase, the temporal profile of the gesture is recorded and used to create a left-to-right Hidden Markov Model by directly associating each sampled point to a state of the HMM. Each state i corresponds to a sample in the training data and is associated with a gaussian probability distribution b_i, which is used to compute the probability of an observation O:

$$b_i(O) = \frac{1}{\sigma_i\sqrt{2\pi}} exp[-0.5(\frac{O-\mu_i}{\sigma_i})^2] \tag{3.1}$$

where μ_i is the *ith* sampled value associated with state i, and σ_i is parameter that can be interpreted as the standard deviation occurring between recorded references and performances. Since the HMM is trained using a single example, σ_i cannot be estimated and therefore must be set using prior knowledge or dynamically adapted depending on the accuracy of the performance. This parameter σ_i is directly related one of the important parameter of the application, that we called *tolerance*, that is defined by $2\sigma_i$ and is expressed in the same unit measure as the sensor data.

Because the states correspond to frames in the original gesture, transitions between states correspond to transitions from one frame of the original motion to another. Three non-zero transitions probabilities occur: a_0, which is the probability of staying in the same frame; a_1, which is the probability of moving to the next frame; and a_2 which is the probability of jumping to two frames ahead. These transitions probabilities correspond to different speeds of performing the gesture: respectively movements that are slower than the original; the same speed; and faster. In order not to bias the model toward certain movement speeds, these transitions probabilities are set to have equal values: 0.33, 0.34 and 0.33, respectively.

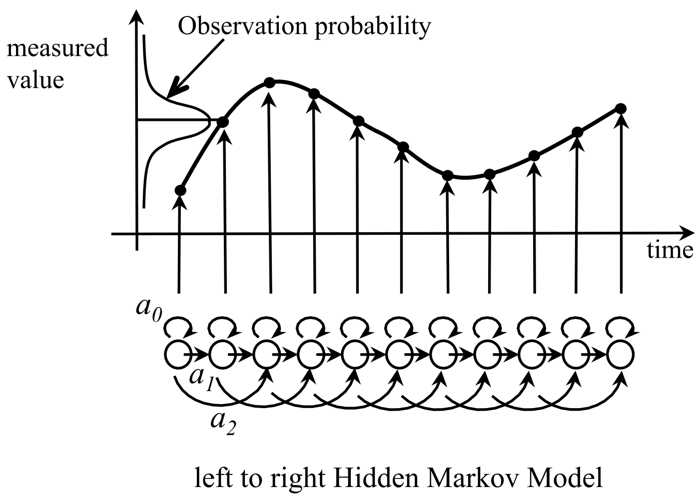

Figure 3.2: Learning procedure: a left-to-right HMM is used to model the recorded reference. The HMM has a separate state for each sample of the training data.

The decoding phase follows standard forward procedure to HMM [Rabiner, 1989b], corresponding to a causal inference (i.e. the inference is estimate without the knowledge of future events, as appropriate standard Viterbi algorithm that operates, without causality constraints, on complete gestures). This procedure requires the computation of a distribution $\alpha_i(t)$ which corresponds to the probability distribution of the partial observation sequence until time t, and state i. This distribution is estimated iteratively in real-time each time a new observation is received and makes it possible to compute two important values: the time progression of the sequence, that is related to the recorded example, and its likelihood. For details regarding such a procedure, please refer to annex A.

Outputs of the algorithm The likelihood estimation depends on the *tolerance* and a second parameter called *latency*. Precisely, for every incoming sensor observation, the system computes a likelihood relative to each reference gesture. These likelihoods are computed by averaging "instantaneous" likelihoods, refereed to each coming observation. The average is computed using a sliding window, which size depends to the number of frames taken into account. For example a window size of 50 frames at a frame period of 20ms will consider one second of the performance. High value of this parameter guarantee more stable results, but it will also add latency to the system in outputting accurate recognition estimation, typically during the transition between two gestures.

Finally, note that the computation of the selection of the correct gesture can be performed in two different manners: either selecting directly the one with highest likelihood value computed as explained above, or adding a constraint on the speed of the gesture performance to be in a given range, such as between half and twice the speed of the reference gesture and rejecting those outside this criteria.

Parameters In gesture recognition, and more broadly in machine learning problems, a good parameters setting is critical to performances. However, as previous studies have shown [Fiebrink et al., 2011], users often have difficulties understanding how to tune parameters of machine learning algorithms properly. One of the goals of this work is to propose and evaluate techniques to make this process as comprehensible and effective as possible, so that users not only can

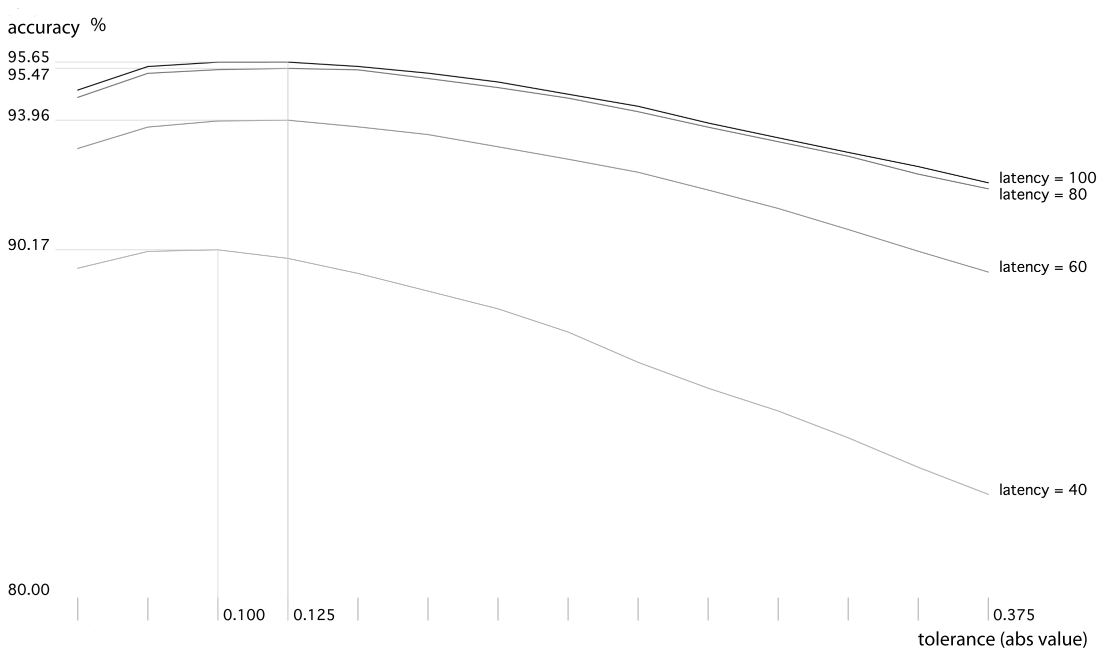

Figure 3.3: Evaluation of algorithm performances in success rate using Wobbrock's 2D database. Results are shown for different values of the "latency" parameter. X axis represents variations of the *tolerance* parameter, which is a parameter of the algorithm explained in section 3.2.3. Y axis displays the success rate. The value of the *tolerance* parameter is expressed in the same unit as the sensor data, which in our evaluation have been normalised by their maximum value so to be in the range of [0,1].

easily understand the meaning of the parameters but can also set them so to *adapt* the behaviour of the system for any specific need they will encounter in their practices.

Figure 3.3 shows how the *tolerance* and *latency* parameters are important and strongly affect the performance of the system. As we can see from the figure, performances of the system against the database converge to an highest peak with a tolerance value equal to 0.125 and a window size equal to 100% of the gesture size used for testing. This graph shows that an optimal setting exists and it is important to guide users into the process of parameters tuning. The next section 3.3, 'Workflow' will explain how the GIDE application provides realtime feedback about the influence of these parameters.

Features We can now finally list the features of the algorithm into the three following categories, which directly related to our desiderata listen in 3.2.2:

Real-time recognition This algorithm returns a real-time moment-by-moment probability that the gesture being performed is the same as each of the pre-recorded gestures in the recorded gesture vocabulary. This probability information is updated continuously while the gesture is being performed. In other words it is updated with a frequency that corresponds to the sample-rate of the incoming sensors signal (typically around 5-20 ms) from the very first sample of the gesture.

Following Our algorithm also tracks a best estimate of the temporal position of the currently performed gesture compared to pre-recorded ones. In other words, in realtime the system aligns users' performances to their gesture references. I refer to this property as *following* a gesture.

Quick learning As explained above, only one example per gesture is needed. This makes the procedure of defining new gestures quick and simple. Section 3.3 will explain in details how this feature is used in GIDE to help enabling an *interactive machine learning* process.

3.2.4 Algorithm evaluation

In order to evaluate the performances of the algorithm, we used the 2D gesture database provided by Wobbrock et al. [Wobbrock et al., 2009] recorded using a Microsoft Surface prototype measuring 24" x 18" set at 1024x768 resolution. This database contains data from 10 users drawing 16 different symbols in two dimensions. Users repeated the drawing 10 times at three different speed rates: fast, medium and slow. The total number of examples within the database are then 10*16*10*3 = 4800. Unlike Woobrock's evaluation, which is offline, we evaluated our system under real-time constraints, without any data transformation that typically require knowledge about the entire gesture, such as the average scaling or rotation angle around the centroid. The only pre-processing treatment we used was the translation of gestures to the origin, which is obtained by subtracting each of the points from the previous ones and can thus be computed in realtime conditions.

For each speed rate and for each user, we iterated over the 10 examples provided considering one series of recording as training-set and the other 9*16 for testing. In Table I we report the success rate of our algorithm after respectively 25%, 50%, 75% and 100% of the gesture length. The table also reports results of our algorithm using the speed constraint mentioned above. We then compare these results with the standard Dynamic Time Warping (DTW) algorithm.

As we can see in table I , results show that Gesture Follower recognition can be almost as good as a standard offline output such as the one provided by DTW and that the correct answer is estimated correctly in real-time in most cases. Our algorithm estimates the result correctly with 62.4% of success rate after a quarter of the gesture, 83.7% at half, 92.2% at three quarters and 95.3% at the end. This also shows that, as expected, the recognition rate increases with the degree of gesture completion. Interestingly, the convergence is relatively fast considering that the difference of the recognition rate between 50% and 100% is only 11%. Note also that the algorithm reaches 97.4% of success rate if an additional constraint is added (which is slightly higher than the 97.1% given by DTW). This constraint corresponds to taking into account only gestures with duration comprised between half and twice the template duration. This is equivalent to taking into account a gesture only if their average relative speed (to the template speed) is between 0.5 and 2.

Table I. Gesture Follower vs Dynamic Time Warping

Algorithm	**Gesture Length in %**	**Success rate in %**
GF	25	62.3953
GF	50	83.6696
GF	75	92.2431
GF	100	95.3385
GF with constraint on the speed	100	97.3675
DTW	100	97.1788

Success rate of Gesture Follower and Dynamic Time Warping algorithms at different temporal position of the gesture length using Wobbrock's 2D database.

Parameters setting is critical to achieve high performances with our algorithm. In figure 3.3 we report performance measurements using Wobbrock's database varying the latency and the tolerance parameters. As clearly shown in the figure, performance results converge to optimal with a given setting of these two parameters.

3.2.5 Implications for the proposed desiderata

Now that the features of the Gesture Follower algorithm have been described, we want to discuss how these link with our proposed desiderata described in 3.2.2.

Continuous control The real-time nature of the algorithm and its moment-by-moment computation of the *gesture recognition* and *gesture following* tasks provides information that can be used to control the target application continuously over time. We refer to 3.3.4 to develop this further by showing three example applications that exploit these features to control digital media in real-time through gesture.

Continuous feedback We take full advantage of the real-time nature of our algorithm to provide *meaningful feedback* to users relating to the output from the real-time *recognition* and *following* aspects of our algorithm mentioned above. This feedback happens in different ways as follows. First, we record video and audio of users when they record their gestures and we align these information streams with the performance in real time as it is happening. This allows users to be able to test their gesture vocabulary seeing and hearing the playback of their recorded gestures as they are synchronised with the performance: it slows down and speeds up exactly matching when users slow down and speed up. The attempt at constant alignment between the performance and the pre-recorded gesture enables users to practice the performance of their gestures when rehearsing (which may involve recording of new improved gestures in the vocabulary.) Moreover, for a more precise comparison between current and recorded performance of gesture, GIDE allows users to visualise how the system is aligning the various streams of sensor data of the performance to the corresponding streams of the performances of the pre-recorded gesture vocabulary. This visualisation offers not only a detailed measure of the differences between the performance and the reference gestures, it is also provides a clear understanding of how the system is behaving in response to the current performance.

User-defined gesture vocabulary and parameter tuning GIDE deeply adapts the way it works for the context it is embedded in by letting users defining the gestural templates it has to recognise. Though this last is a property which is shared in many common systems, as we discussed in section 2.2.3 these often require users to build a large dataset, which is a difficult and delicate task. Conversely, as explained above, the algorithm used by GIDE allows users to define their gestures by recording them just once. Moreover, GIDE supplies guidance in tuning the most important parameters of the system by providing a corresponding graphical feedback on data visualisation and, also, by using metaphorical names for these parameters. This enables users to personalise the system behaviour with increased precision in intuitive ways. The act of recording gestures, testing them and tuning system parameters are proposed as three tightly interleaved processes that make the gesture design workflow an interactive and fluid process. In this way, dancers can define gestures through performing dance, musicians can define their gestures through performing music, and players of computer games can define gestures through interacting with the game.

We will show how the combinations of the features described above allows the design of the application to be particularly *easy to use* by non-expert users. The realtime nature of our system, and consequently the continuous feedback provided by the application, allows users to have a clear understanding of how the system responds to their actions. Furthermore, the graphical feedback on the effect that system parameters have on the performances of the system, combined with the adoption of metaphorical names, aims to make the task of parameter tuning (which has been traditionally difficult in machine learning) as easy as possible. In Section 3.4, we provide details about the usability of the application through an user evaluation case study.

3.3 Workflow

In order to make the usage with the algorithm as clear and straight forward as possible, GIDE proposed a three states interaction paradigm (figure 3.4). These three phases are tightly interleaved and allow user interaction with all aspects of the design process: gesture recording, gesture following and parameter tuning. To accomplish the *fast and focused* UI principle, the entire process is iterative and users can quickly switch between phases during the design process. In this section we describe these phases in detail.

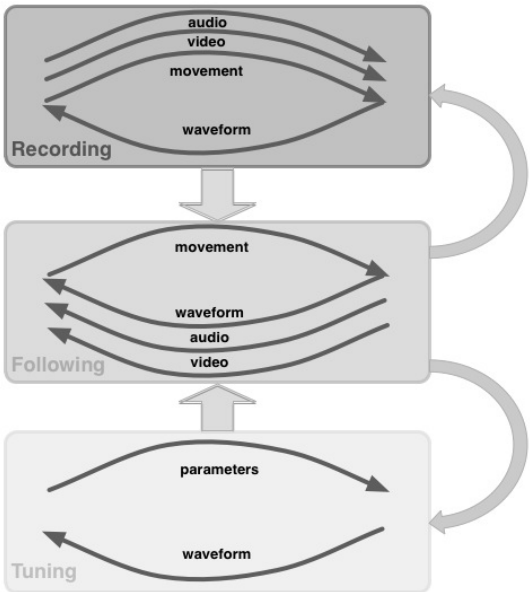

Figure 3.4: *GIDE workflow.* The gesture design workflow divided into three tightly interleaved phases: gesture recording, gesture following and parameter tuning.

3.3.1 Phase 1: recording a gesture

GIDE allows users to easily build the gesture vocabulary. They can quickly record a gesture, view it in the graphical interface, test it and, in the case they are not satisfied with it, record it again. We refer to the set of recorded gestures as *gesture vocabulary.*

Each gesture within the vocabulary is accompanied with a graphic component, called the "gesture editor" as seen in Figure 3.1. A gesture can be of any length of time and it is represented by the following components: a name, a multi-waveform for sensor data and optionally a video sequence and an audio waveform. As we described in the previous section, our algorithm allows users to define gestures by recording them just once. This ability has allowed us to design this phase to be as close as possible to the one of recording audio and video in standard AV production sequencers so making the handling of sensors data as straightforward as possible.

At the beginning of a new session, the gesture vocabulary is empty. The user can then decide to record a gesture either by pressing the "record button" via the GUI or triggering the 'record' command remotely. In the Evaluation section, later in this chapter, participants were able to trigger the record function both with the mouse and with a Nintendo Wiimote.

Once this happens, the application starts recording incoming sensor data together with video and audio from attached cameras and microphones. Users are encouraged to also record a sound while recording a gesture (for example spoken sentences), in order to have richer feedback during recognition. This multimodal stream of data is graphically represented in the *Input View* on the top-left on the application. Furthermore, during the recording the user can see the recorded data in the *gesture editor* related to the gesture that is being recorded.

After the recording, the user can select a part of the gesture (for example to discard a silence at the beginning or the end), zoom, scroll and playback at different speeds. A button called "Pop-Out" is available to open a new resizable window if more space is required. It is also possible to add a *temporal marker* in a specific point of the multi-waveform by double-clicking on it.

3.3.2 Phase 2: "Follow" mode and real-time feedback

When the gesture vocabulary contains at least one gesture, users can start evaluating their vocabulary with a real performance. GIDE supports the traditional batch testing, present in classic machine learning tools, but also a realtime testing called *follow* mode. Batch testing is described in more details in the next section.

In *follow* mode, as users perform a live gesture, the application gives a moment by moment probability estimate of which gesture they are performing and *where*

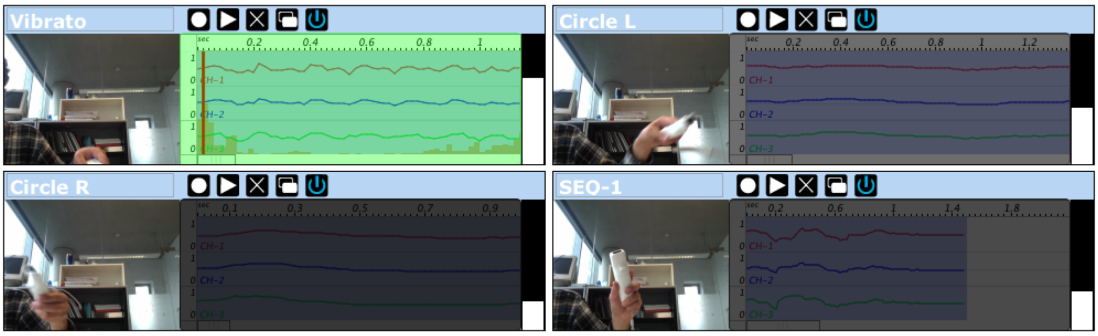

Figure 3.5: *Realtime gesture recognition.* GIDE allows for realtime gesture recognition continuously computing a likelihood measure between the performance and every pre-recorded gesture. The likelihood of a gesture graphically corresponds to a level of transparency. The green gesture is the likeliest one. The *contrast* parameter increases or decreases the difference between high and low likelihood values, so the associated colours.

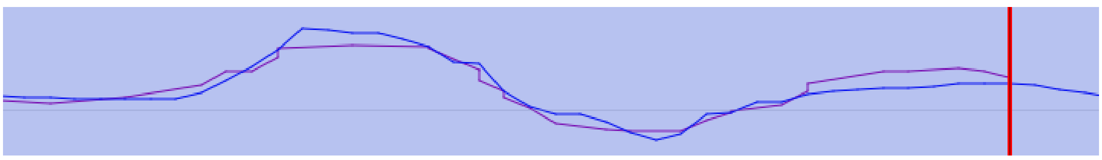

Figure 3.6: *Realtime gesture following.* GIDE allows for realtime gesture following aligning the performance with pre-recorded gesture continuously over time. In the figure we see the waveform of the incoming data stream (purple) aligned frame by frame with the corresponding position of the pre-recorded one (blue). The red cursor represents the temporal position of the gesture.

they are within that gesture. The system performs continuous recognition based on incoming data and gives a realtime estimation on its similarities against each pre-recorded gesture.

The probability of each gesture is represented visually by how transparent the associated editor is, while the likeliest one becomes green (Figure 3.5). As well as the current probabilities for each gesture, we have a precise estimate of the temporal position within the gesture. As previously mentioned, we refer to this feature as *following* a gesture.

This enables GIDE to provide a realtime multimodal feedback on the recognition during the performance. This multimodal feedback is composed by the three following aspects:

Video alignment Each video panel plays back the pre-recorded video synchronised with the performance. Typically this allows users to compare their performance with the corresponding video image of themselves when performing the recorded gesture.

Audio alignment The audio that users recorded is played back synchronously during the performance. The application allows the user to decide between playing back only the audio of the likeliest gesture or to do a *mix*, i.e. associating each gesture to a volume playback that is proportional to its likelihood. In this way users have an auditory feedback on which gesture has been recognised. They also have an auditory feedback of the alignment of their performance with the pre-recorded gesture, as the pre-recorded sound is played back on the temporal position of the *follower*.

Waveform alignment The temporal position of the performance within the pre-recorded gesture is displayed through a red cursor over the data multi-waveform (Figure 3.6) together with a *probability function*, an orange waveform that displays the probability associated to every frame. Furthermore, we have implemented what we call an "alignment view" which, when enabled by the user, displays a pink multi-waveform superposed to the original one, representing the incoming sensor data aligned to the reference gesture frame by frame. In this way, users can clearly see the difference between their performance and the pre-recorded one as a vertical distance between the two multi-waveforms in every point. This representation works particularly well in association with the *tolerance* parameter described later in the chapter.

3.3.2.1 Batch testing

Within phase 2 GIDE also supports batch testing by a facility called the *testing performance*. In our design we have given the testing performance a very similar appearance to the gestures contained in the *gesture vocabulary*. The *testing performance* allows the user to record an arbitrarily long *real-world dataset* and then test it iteratively against the *gesture vocabulary* while changing gestures and parameters to obtain best results.

When the user clicks the *test* button, the system reads in a row all the data stored into the *testing performance* as if this data was coming in real time from a performance by a user and so instantly highlights all the areas in its multi-waveform where the likeliest gesture reached a *threshold* given by the user.

As for the gesture vocabulary, it contains both sensor data, video and audio and supports the *retrospection* property: users can select gestures, play them back and re-record.

Thanks to the low computational cost of the algorithm, the time for testing a dataset of few minutes is typically few milliseconds. This provides the user with information in order to redefine the gesture vocabulary and tune parameters in an interactive way, seeing the results appear instantaneously in the graphical interface.

3.3.3 Phase 3: parameters tuning

As previous studies have shown [Fiebrink et al., 2011], users often have difficulties understanding how to tune parameters of machine learning algorithms. However, as we show in figure 3.3, parameters of our algorithm are critical to reaching high performance in the recognition task. We provide support for this process in three different ways.

First of all, we assigned a name for each parameter that aims to supply a useful *metaphor* for the user describing a common digital media practice. Second, we have added short *text hints* about how to use each of the parameters. Finally, the effect of two of the three parameters (tolerance and contrast) have a corresponding *representation* in the graphical interface, helping users to better understand how they affect system performances.

1. *Tolerance*

   The role of this parameter in the algorithm corresponds to a constant standard deviation of the Hidden Markov Model as has been explained in section 3.2.3. This parameter has been explained to users as *"how much the performance is allowed to be different from pre-recorded gestures"* and we have therefore named it *tolerance*. This name is a useful example of a metaphor based on common digital media practice: in Adobe Photoshop there is a parameter with the same name which determines the range of colour that the Magic Hand tool selects. Similarly, in GIDE this parameter is graphically associated with the *thickness* of the sensor data multi-waveform and shows the range of values that determine whether the performance belongs to the gesture. We have found that this works particularly well in combination with the *"waveform alignment"* described in previous section, as users can see the distance of their aligned gesture compared to the 'tolerated' range of values.

2. *Latency*

   The actual probability that each gesture recorded in the vocabulary is a match for the current performance is computed as an average of probabilities calculated for each frame and stored on a sliding window. Thus the size of this window specifies the amount of time taken into account for accurate gesture estimations. The effect of this parameter is basically to affect the latency of the system. If the parameter is set high, it will recognise gestures highly reliably but it will react slowly to changes in user input. If it is set low the system will react faster but less reliably.

3. *Contrast*

   In our system, the value of the probability of each gesture in the vocabulary is normalised such that their sum is always equal to 1. For practical reasons, we have designed a parameter to tune this normalisation in order to increase or decrease the difference between high and low probability values, following the formula $l(i) = e^{e^{c*l(i)}}$, where $l(i)$ is the probability associated to the gesture i and c is the contrast value.

The definition that is given to the users is: *"turning up the contrast parameter heightens the differences between gestures."*

The word *contrast* works as a metaphor as we think about the contrast of an image quickly. As gesture probabilities are graphically represented as the transparency of their associated gesture editor, tuning up the *contrast* parameter will increase the contrast of the colours of such editors. Figure 3.5 shows the application with the *contrast* parameter set to a high value.

3.3.4 Three DMIs based on GIDE

Having looked at the workflow of the application for designing new gestures, we now move our attention to consider how user-defined gesture following can be applied in real-world scenarios. Here we show three different standalone applications for the gestural control of digital media. Theses scenarios are based from cases that were previously prototyped with the *gesture follower*, but without the integration of the user interface of GIDE.

1. *Video scrubbing* This first application is inspired from used in the installation *if/then installed* (by Siegal, Bevilacqua, Berenger, Goidell and Lambert: `http://www.thebakery.org/interactive-if-then-installed`). This installation, using the *gesture follower* algorithm, demonstrated the interest of a "videoscrubbing", which is explained below. The installation was designed using pre-defined gestures. The use of the GIDE interface, allowing users to add their own gestures, could extend this interaction paradigms to a wide range of applications.

   The gesture-driven video scrubbing works as follows. First, users select different video files, one for each gesture they want to learn. Then, as soon as they start recording a new gesture, the corresponding video file is played back. This allows users to 'mime' to the video while they record their gesture. When users switch to *follow mode*, the video of each of the recorded gestures in the vocabulary is aligned to the most likely position and played back by GIDE. The user can switch between the *likeliest* mode, where only the video that corresponds to the likeliest gesture is played, and the *mix* mode, where all videos are played back and their transparency is mapped with the likelihood of the associated gesture. This could result as a gestural interface for VJing, where users can continuously control the video playback though their gestures.

2. *Supervised continuous sonification*

   The second application is similar to the one described above but uses sound instead of video. It allows the user to continuously align the playback of a sound file with a gestural performance. This paradigm was previously validated in pedagogical scenarios where students can "conduct" recorded music using gesture input [Bevilacqua et al., 2010].

   First, the user loads different sound files, each one associated with an empty slot in the gesture vocabulary. Second, as soon as the user starts recording a gesture, the loaded sound associated to it is played back. Thus the user listens the sound when recording the gesture, adapting the performance with the tempo of the sound or *mime* the sound itself.

   When the user switches to *follow mode*, the sound is played back following the temporal position of the performance, which is given by GIDE. So, when the speed of the performance slows down the sound playback slows down as well; when the performance accelerates, the playback accelerates as well. The sound is thus continuously *aligned* with the performance. This allows the user for a *supervised sonification* of a gesture based on previous recording.

   In order to enhance the quality of the time stretching, we employ a phase vocoder. This technique allows to leave the pitch of the sound unchanged while changing its playback speed. As for the previous application, two options are available: in the *likeliest* mode, only the sound associated to the likeliest gesture is played back; in the *mix* mode, all the sounds are played back and their volume corresponds to their likelihood.

3. *Triggering*

   The last application allows to trigger a series of digital media based on discrete temporal positions along a single gesture. Such scenario is similar to previous artistic applications of the *gesture follower* allowing gesture-based system to trigger sound processes, as described in [Bevilacqua et al., 2012].

We have designed GIDE so that it is possible to place a named marker at a specific temporal position within a gesture. During a performance, at anytime the temporal position of the likeliest gesture reaches one of these temporal markers, a different sound of video can be played. Other types of digital media, such as MIDI notes and light control, could be controllable in the same way.

Having described the interaction design process and looked at the architecture of the application in detail and especially its novel mechanisms for real time feedback, we now move onto a presentation of our set of experiments to show how the system was used in practice.

3.4 User evaluation

In order to evaluate how the features of our system, as defined in subsection 3.2.2, are practically used by end users, we performed a user study with 23 participants among artists, musicians and designers. For the sake of clarity, we summarise the purposes of our study in the list below.

- to measure participants' degree of interest in the possibility of being able to *control and manipulate* digital media continuously over time through gestures;
- measure whether the various *types of feedbacks* our system provided were effective in helping to control the system, as well as assessing the individual importance of each feedback component;
- measure how well participants were performing in the process of customisation of the system by *defining their own gesture vocabulary* as well as evaluating whether this process was pleasant or frustrating for them;
- measure the effectiveness of *parameter tuning* and whether the feedback was accurate enough to guide them in this process to obtain best results;
- studying the different *strategies* participants adopted in their practices in order to define their gesture vocabulary and how such strategies developed during the workshop to possibly obtain better results.

3.4.1 Participants

The workshop was organised by inviting 23 participants from different domains including five professional music players, thirteen electronic music performers, three visual or interactive artists, one dancer-choreographer and one programmer. All participants had experience of using digital media in their artistic practice and were familiar with many standard digital production tools such as Digidesign ProTools and Apple Final Cut. In addition, 15 of the 23 were familiar with visual programming environments such as MaxMSP and PureData, 7 of these were also used textual programming language in their artistic practice, and 4 of these had some familiarity with machine learning theory. 8 of the 23 had no familiarity with anything other than standard digital media production tools.

Ten of our participants were PhD students in music and computational art, two were university faculty members in music and the others were independent artists. All participants were aged between 23 and 35 and the experiments took place over six sessions in London and Edinburgh. Each session had between 3 and 8 participants and lasted about two and a half hours. We named the workshop “Workshop on realtime gesture recognition for performing arts” and all participants applied spontaneously and were not remunerated.

3.4.2 Workshop procedure

Every session of the workshop has been divided into five parts and participants were asked to fill in the relative section of our questionnaire after each part. All sessions were video recorded in their entirety. The five parts are introduced below.

1. **Introduction to gesture recognition systems**

   We started each workshop by explaining to users the general concept behind gesture recognition and the difference between direct mapping from sensors data and control parameters as opposed to user-defined gesture recognition.

   The first video depicts the *video scrubbing* paradigm previously described and can be found at http://tiny.cc/9mpibw. It detailed an interactive installation that was built several years ago (using the previous *gesture follower* system) and shows a dancer performing live in front of a big screen. The screen displays a recorded second dancer that appears to mirror the live dancer doing the same actions at the same time. This video clearly shows an interaction based on a continuous output of the gesture recognition system.

   The second example, called “Augmented Violin”, explains the *supervised continuous sonification* paradigm (also using the previous *gesture follower* system). It shows a live violin player performing a piece at different speeds while a second recorded violin (recorded by the same player) accompanies the live performance following the tempo.

   The section of the questionnaire relative to this first part asked to users to evaluate whether they felt they understood the basic concept of continuous gesture recognition and what was the interaction between the gesture and the sound in the two videos we showed.

2. **Playing with an existing application**

   We showed to users a video about an interactive installation called *Granularia* that we presented at the Festival of Science in Genoa in 2010 (which is publicly available at `http://tiny.cc/d1bux`). This installation allowed the user to control different sound engines by moving a mobile phone on the air. Specific gestures were recognised and used to trigger associated sounds. We then ask users to try the same application using a Nintendo Wii Remote in order to get familiar with the possibilities offered by supervised gesture interaction. Through the questionnaire we asked whether users understood the goal of this application and were able to control this system reliably.

3. **Designing a single gesture**

   At this point we asked users to run the GIDE application on their computers. We demonstrated how to record and follow a gesture through the application using a Nintendo Wiiremote. We explained that both video and audio were being recorded, and showed a basic example of an association between a gesture and a vocal sound. We then asked users to try to record their own gesture in their computers and evaluate the different components of the application.

4. **Building a gesture vocabulary and testing the recognition in realtime**

   In this phase users were asked to record several gestures to create a vocabulary and experiment whether or not they could be triggered in a subsequent performance. They tested the gestures that they had recorded switching the application in the 'follow mode' and performed similar gestures again looking at the various realtime feedbacks provides by the application as explained in section 3.3.2. We also gave particular hints about how to tune parameters as explain in section 3.3.3. As they were perform this task, we recorded and monitored the strategies they took to record their gestures and tune the proposed parameters.

5. **Batch testing** In this part of the workshop we explained to users how to evaluate performances using the batch testing feature we described in section 3.3.2.1.

6. **Developing an application for gesture sonification**

   As users got familiar enough with the application, we showed them how to use information provided by GIDE to accomplish the task explained in section 2 which is supervised continuous gesture sonification. We showed how to create an application in the MaxMSP environment `http://www.cycling74.com` that loads sound files and develop a mimicking paradigm between gestures and sounds through gesture recognition: the provided application allow users listening to a sound and mime it with their hands; then perform the same gesture again at a different speed and hear the stretched sound.

7. **Developing an application for video scrubbing**

   In this part, we explained the video scrubbing paradigm (section 1) and showed users how to build their own application using the MaxMSP environment.

3.4.3 Measures

We evaluated the study through both a questionnaire and a semi-structured interview. After each of the seven sections of the workshop just described, we asked participants to fill the relevant part of the questionnaire. For all but the first two sections of the workshop we asked users if the application worked as they expected to accomplish the proposed task. Furthermore, we ask them to evaluate the different components of the application, which are the video scrubber, the waveform cursor, the alignment view, the probability waveform, the tolerance parameter, the responsiveness contrast and the background colour changing. Questions about both usability and evaluation were repeated for each section of the workshop with each question presented as a 7 point Likert like scale. Users were invited to add commentaries at every stage. The semi-structured interview took place after the session and asked how much participants were happy with their results, the kind of strategies they used to design their gestures, the level of usability and usefulness of the system and whether they would use this application for their works.

3.4.4 Results

In this section we first present the results of the questionnaire and we relate them to the achievement of our proposed desiderata. We then discuss the different strategies used by participants. Finally we debate some issue of the application arisen during the workshop.

3.4.4.1 Questionnaire

The questionnaire results are shown in table II (3.7) . The questionnaire responses were analysed with a one sample Wilcoxon signed-rank measuring the difference between the sample responses and the mid point of the scale (4). The mean of all answers was above the mid point (indicating a favourable response) with all but two being significant to at least p=0.1 and the majority being significant to p=0.001. Some of the later questions had a lower number of responses due to not all participants reaching the last part of the study in the allocated time. This might account for the lower levels of significance to the later questions. For all sections participants were asked whether the system worked as they expected, in all sections the mean answer was higher than the midpoint with p=0.01, except the final (video scrubbing) section where the significance was p=0.05. For the final two sections they were asked whether they understood the interaction between their gestures and the sound or video, mean answers to both these questions were significantly above the midpoint to p=0.01. In the final two sections they were also asked whether they could control the system reliably, mean answers to these questions were significantly about the midpoint to p=0.05 (audio condition) and p=0.1 (video condition). Participants were also asked to rate the usefulness of each of visualisations for each stage of the study. Mean ratings were significantly above the midpoint in all but two cases. There was no significant difference in the ratings of different visualisations.

3.4.4.2 Achievement of the properties

In this section we describe participants' behaviour at each stage of the study in respect to the desiderata of gesture interfaces we defined in section 3.2.2. Although the questionnaire responses have been substantially above threshold, this is generally not a sufficient factor to justify users engagement and understanding as it is often biased by the tendency of participants to associate high results during a study. It is however useful to compare the results of the questionnaire regarding each components of the application against each other and integrate the findings with the observations annotated during the study.

1. **Continuous control**

   Continuous control has shown to be a very important feature of GIDE in terms of the range of applications it can allow to control. In section 3.2.3 we have shown that our algorithm is capable of real time control and our participants answers to the questions "Did you understand the interaction between the gesture and the [sound/video]?" and "Are you able to control the system reliably" shows that they were able to understand and use continuous control. Furthermore, in the semi-structured interview at the end of the workshop, when we asked participants to imagine how GIDE could be useful for their own practice they answered enthusiastically and many of the strategies they developed, described in section 3.4.4.5 are clearly inspired by the possibility of controlling the target application continuously over time.

2. **Continuous feedback** The continuous control features of the Gesture Follower algorithm also enable us to give the different forms of realtime multimodal feedback described in section 3.3. This ensemble of visualisation and sonification techniques aims to give a clear snapshot about the state of the algorithm at every instant of a gesture performance. A comparison between these different techniques is provided in 3.4.4.3.

3. **Definition of personal gesture vocabularies and parameter tuning**

   All of our participants were able to record gestures of their own design, use those gestures coherently with their intentions and obtained the expected results, demonstrating their ability to personalising the application by tuning parameters of the algorithm and using different strategies for designing gestures. We discuss these points in details in sections 3.4.4.4 and 3.4.4.5.

3.4.4.3 Comparison between the different types of feedback proposed

When been asked to compare the different components of the GIDE application, as a first reaction 17 participants of the study claimed that they could not decide as it was the combination of all of them together that was needed to be useful. Some participant even remarked that only one of them on its own would not be enough for a controlled performance. However, deepening into the questions, important differences arose regarding why and when they focused their attention on each individual components.

Video resulted to be the most important components at the beginning of the study, but gradually less important compared to the other components as users got more familiar with the application. The main motivation to use the video feedback has been its effectiveness in helping to remember the individual gestures, reported by 9 participants. Amongst them, the choreographer and another one also imagined that if they did not work on a piece for several days they would then need to look back at the video to remember how they performed as the waveform alone would not be sufficient.

Participants described the *sensor multi-waveform* as a way to get an overview of the whole gesture. Almost everyone also focused the attention on the red *cursor* over the waveform in order to understand the concept of following a gesture. With the exception of the computer artists, participants where not familiar with the concept of acceleration data. Amongst them, 4 reported a certain difficulty in understanding the data represented in the waveform and confessed that they confused it with absolute position, expecting higher values when the sensor was higher. 3 participants familiar with audio-video production tools found an analogy between audio waveforms and the way the sensor waveform was displayed in GIDE. Others participants managed to get familiar with it during the workshop, although 6 also reported that they were not able to discriminate between the 3 different axes of the acceleration and they would focus more on the global 'motion' of the gesture. This confirms results that observe how raw motion data is often not expressive enough and we need a better representation (see for example

[Kratz and Ballagas, 2009] and [Linjama et al., 2008]).

Every participant recorded their voice in association with the gesture, mostly because this is what they saw myself doing when explaining the system. Most participants associated specific words or sounds to particular parts of a gesture instead of recording one continuous sound. For example, different syllables or screams were often associated with the more energetic parts of the gesture. 3 of them recorded sounds other than their voice, such as hand claps, hitting the floor with their feet or tapping the table with their hands.

Auditory feedback was felt to be the most precise form of feedback in association to individual gestures' likelihood and to the speed of their performance. 5 participants (3 music players, a visual artist and the choreographer) spontaneously remarked their interest in using auditory feedback as it permitted them not to have to be forced to look at the screen, thus enabling eye-free performances. However, when asked they stated that this was possible only after a number of iterations with the system, as this freedom strongly relied on the ability to remember their gesture vocabulary quite well.

3.4.4.4 Parameter tuning

In general, users tuned parameters quite often when defining their gestures and when asked *When you were not satisfied with a gesture, did you prefer to record it again or find a better parameters setting?*', they all claimed they did both. A user said: "*if the system is kind of working but not very well, I try to play with parameters, but when it doesn't work at all I preferred to re-record again*". All other participants of that session agreed with him and we saw this behaviour across the workshops in general.

During the first task of the workshop, when we asked participants to record only one gesture and *follow* it, initially they usually re-performed the same gesture either in the same way or slower and judged the result based on the auditory feedback and the red cursor. Often, when they tried to perform something that was too different from the pre-recorded gesture, they understood that the system was not following the gesture very well by seeing the red cursor suddendly 'jumping' to different temporal positions and by receiving a noisy auditory feedback. In those cases they were able to understand the problem quite quickly and work to find a solution, either recording the gesture again or changing the *tolerance* parameter. We gave them practical tips about this last parameter, such as "*if the cursor starts going forward by itself, it means that the system is too tolerant*", or "*if the cursor starts jumping too much and the audio starts becoming mad even if you are performing well, turn up the tolerance a bit*". The association between the tolerance and the thickness of the waveform was straightforward to understand for all our participants.

When we asked users to record more than one gesture to experiment with recognition, they all started recording very basic gestures each very different from each other. This allowed them to make the system work immediately and get a real sense of its behaviour, before moving on to record more complex and subtle gestures.

Using an accelerometer as an input device, some of them initially recorded gestures in which the motion in one axis was clearly predominant compared to the others which was useful because it often enabled them to achieve a quicker and better understanding of the meaning of the waveform.

The *contrast* parameter was used mainly when the system was uncertain between two or more gestures. By increasing this parameter they could see the likeliest gesture more clearly referring to it as the *green gesture*, pointing out that the association between the likelihood and this specific colour was pretty clear. On the other hand, when colours started flickering too much, they knew quickly that it meant that it was a good idea to decrease the value of the contrast.

One of the more surprising results for us was that the *latency* parameter was much less used than the others. Some participants admitted that the effect of this parameter was not clear and reflecting on it at the time we thought it was because there was no graphical feedback associated with this parameter which provoked a fear in the user to changing the value of this parameter. In response we will release in the next version of GIDE a graphical feedback to this parameter, i.e. highlighting the part of the data input view that corresponds to the amount of time specified by the parameter.

3.4.4.5 Observed users strategies

GIDE proved to be a tool that engaged the participants of the workshop and that proved to be comprehensible and flexible enough to be used in different ways and for different applications. The heterogeneous background of the participants has been reflected in the strategies adopted and the potential usages they imagined. We discuss the three most evident ones here.

Both a *professional piano player* and a *professional choreographer* spent most of their time training themselves to perform their own gestures reliably by recording quite expressive gestures and using their voice as audio at the same time. They both used the *warping view* function to measure the differences between performance and that which they pre-recorded gesture and were not satisfied until the audio output was sufficiently close to what they recorded. It was interesting to see how they played with the value of the *tolerance* parameter quite a lot. The choreographer said that they had a much better concept of accelerometers after trying to perform the same gesture several times and watching at the *warping view*.

The choreographer particularly liked the *testing performance* component. She said that the user of the application and the performer would often not be the same person and this enabled her to record the performance just once and work on parameter tuning later. She also explained the *following* mode is interesting for live situations, where the choreography and other theatrical and technical affects need to be directly synchronised with the performer and so not requiring a human to trigger them. She also suggested a new feature that we had not thought of previously. Her idea was to build a gesture vocabulary *from* the testing performance: this means recording the *testing performance* before, and then copy and paste certain parts of it for defining gestures.

Computer artists generally preferred a more methodical approach tending to think about their gestures in advance and then record one gesture after the other. Then they usually tested one single gesture or a fixed series of gestures changing one parameter at time. Other *contemporary music performers* took a radically different strategy and recorded quite complex gestures and then they played with the system trying to 'confuse' it. One user said “I got how the system works, now I just want to hack it”. In doing that they also played a lot with the tolerance and the contrast, keeping the contrast high enough to see major differences between their gestures. This shows that participants with different artistic background developed different strategies for using GIDE that were personalised to their needs.

Finally, it is worth to mention that working with accelerometers caused some problems. First because participants who were not familiar with this kind of device found it simpler to think in terms of absolute position. Furthermore, they were surprised to see that, due to hardware limitations of accelerometers, slow movement was not recognised.

3.5 Discussion

This chapter presented the motivations, design, implementation and evaluation of a gesture interaction system called GIDE. The system leverages the features of the Gesture Follower algorithm in order to provide realtime and continuous analysis about users gestures. The interaction workflow proposed, as well as the ensemble of visualisation and sonification techniques employed, add *visibility* to a complex system in order to ease and enhance the usage of the algorithm and allow the system to be flexible and open enough to allow users to freely explore and personalise different gestures so as to make the system's customisation process so rewarding and straight forward to become embodied in the interaction process itself. I have proposed a set of three properties that define a novel usage of gesture interaction systems which we can now call *fluid gesture interaction design*.

This section summarises the main contributions and limitations of the proposed work as well as outlining future research.

3.5.1 Contributions

The contributions of this chapter can be summarised in three different key points: the evaluation of the Gesture Follower algorithm with its parameters 3.2.4; the evaluation of the interaction workflow described in 3.3 and the comparison of the different feedback components presented (3.4.4.3 and 3.4.4.4); the description of the different users strategies arisen during the workshop (3.4.4.5).

Evaluation of the algorithm Through a quantitative evaluation on a standard set of 2D gestures, we demonstrated that the algorithm used by GIDE can perform as well as that of Dynamic Time Warping (DTW) which is a standard template-based algorithm in the field of gesture recognition systems. Nevertheless, unlike DTW, the algorithm can provide moment by moment information *during* the gesture performance itself, such as the relative speed, and a early estimation of the recognition results. The recognition efficiency of the algorithm depends on the tuning of two parameters called *tolerance* and *latency*. I have demonstrated that the dependance of these parameters to the recognition efficiency follows wide bell-shaped curves with one maximum. What this means is that, whatever the initial value of the parameters, the optimal values can easily be found manually by any user through simple trial and error. The user simply tunes the parameters so as to increase recognition efficiency and when no further increase can be found the user can then be confident they are operating with a system which is operating at maximal recognition efficiency. This ability - for users to easily tune system parameters for maximum efficiency - is an extremely important result in terms of demonstrating how the system can be customised and used by a range of users with different expertise.

Comparison of the different forms of feedback The second aspect of this work concerned a controlled study with 23 users from different performance background in which the proposed interaction workflow has been evaluated and the different features of the application have been compared with each other. The different feedback components presented different and complementary features. The video alignment, which shows the video of the users in the act of recording a given gestures and played back at the speed of the one being performed in realtime, proved to be useful at the beginning of the workflow in order to remember the recorded gestures. The waveform alignment was mostly used to have an overview of the full length of the gesture and to measure the difference between a performance and a specific pre-recorded gesture in details. The audio alignment proved to be the most useful form of feedback when a certain familiarity with the application was achieved and to provide the most accurate feedback about the temporality of the performance. The possibility of not having to look at the screen during the performance through the exclusive use of the audio feedback proved to be interesting and spontaneously noted by 5 of our participants.

Observed users strategies Three different users strategies have been observed during the workshop and described in this chapter. These strategies and modes of interaction are strongly related to the different backgrounds of the participants and the objectives they had in mind whist learning to use the application. For example, the choreographer was used to work with written choreography that the dancers have to follow while performing a given piece. For this reason, having a way to observe in all the details the differences between a performance and the pre-recorded gesture was found to be extremely useful in order to understand the behaviour of the system and compare it with the given choreography.

We saw how quickly participants were able to understand and enjoy using the system, engaging with the real-time gesture *recognition* and gesture *following* during the workflow of gesture design almost immediately. Almost every user taking part in the evaluation was able to control the application reliably after a few minutes and to develop a set of gestures to control the system in a way that they found satisfying with system behaviour meeting intention and expectation. The evaluation demonstrated how users could seamlessly move between recording new gestures, testing them and tuning parameters and that this enabled the relationship with the system to be fluid and spontaneous. Most of all, participants adopted very different working strategies imagining a wide range of possible applications across different domains and were extremely enthusiastic about the future potential for developing new performances in their own creative practice.

The system presented enables users to *design by doing* where gesture interfaces are created by performing gestures. Our participants' enthusiasm supports Fiebrink's conclusion that this *embodied* way of designing gestures is both liberating for users and allows for the creation of rich styles of interaction [Fiebrink et al., 2011]. The fact that GIDE gives real-time feedback about the recognition process in different modalities was found by participants to be extremely useful at different stages in the workshop. They claimed that the combination of (i) video to remember the details of the gesture, (ii) audio for precise and instant feedback, and (iii) data waveform for the ability to see an overview of the gesture over time and its temporal alignment with the pre-recorded gesture was critical to their engagement with the system. This real-time feedback on the performance of gestures also helped participants understanding the function of the various parameters of the algorithm and so to be able to tune these parameters to produce their desired results effectively and efficiently. I believe that this is an especially significant result as the parameters setting has been a challenge for interactive machine learning [Fiebrink, 2010]. If non-expert users are able to effectively tune these parameters as we have described in this chapter, it opens the way to using more sophisticated machine learning algorithms, such as Hidden Markov Models, that require considerable amount of tuning in order to be effective. Moreover, it paves the way for experimenting as to how realtime feedback can become a crucial feature of future interactive machine learning research in general.

3.5.2 Limitations

This chapter attempted to demonstrate the importance of making the process of the personalisation of the GIDE interface as quick and engaging as possible, with users recording gestures just once, in order to make this operation part of the interactive workflow. This enables the whole recognition process easier to understand, as gesture performed by users simply need to be similar to the ones they pre-recorded. This is very different from the standard approach where systems create models of gestures based on a large number of examples which are hidden to the user and so make system much less transparent. However, this method has certain limitations as it does not support generalisation and any performance needs to be sufficiently similar to the set of pre-recorded gestures. While it is important not to loose the feature of "quick recording", adding new system functionality that enables users to record more than one example per gesture, and therefore create a more complex and flexible model of the gesture defined by the user, is very much central to the ongoing research investigation. It can be envisaged the possibility to achieve this is by allowing users to record one or more examples for each type of gesture and then building separate Hidden Markov Models for each recorded example. Then, at run-time, recognition can be achieved by considering the correct type of gesture to be the one which corresponds to the HMM with highest likelihood value. However, this would require the designer to research new methods for data visualisation and auditory feedback, as well as new ways of interacting with the system, but in such a way would still keep the workflow as fluid and intuitive as it currently is.

Another problem of the proposed approach is in handling the output during the very beginning of a gesture, when the likelihood of recognising the correct one is still low (table I). This results in a period of uncertainty when the system is first started and also when a user transitions from making one gesture to another. For certain scenarios, this issue can be handled at the application level. As showed in the two applications described in section 3.3.4, audio and video scrubbing, this feature is used to blend between different media based on the likelihood value of their associated gestures. As GIDE provides continuous likelihood estimates for each gesture, transitions between gestures will corresponds in smooth transitions between different likelihood levels until the algorithm gets a clear result. This was a reasonable solution for the DMI described in the chapter.

However, for different applications where a more defined segmentation is required, further research is needed. For instance, it would be possible to add constraints to avoid that users record gestures that are too similar at the beginning. Another solution could be to add a variable latency to the system to compensate the initial time of incertitude.

Table II. Questionnaire results

QUESTION	N	mean	W	
1 - Introduction				
Did you understand the interaction between the gesture and the sound?	19	6.63	190	****
2 - Playing with an existing application				
Do you understand the goal of the application?	15	6.93	120	****
Are you able to control the system reliably?	15	5.93	105	****
3 - Designing a single gesture				
Did it work as you expected	22	6.00	210	****
video	22	6.05	224	****
waveform cursor	22	5.86	241	****
warping view	22	6.00	231	****
probability waveform	22	5.27	183.5	***
tolerance parameter	21	6.43	231	****
4 - Building a gesture vocabulary and testing the recognition in realtime				
Did it work as you expected	22	5.59	219	****
video	22	5.91	206	****
waveform cursor	22	5.77	204	****
warping view	20	5.10	169.5	**
tolerance	22	6.09	210	****
responsiveness contrast	21	6.05	229	****
background colors	22	5.55	162.5	****
5 - Batch testing				
Did it work as you expected	20	6.30	190	****
video	18	5.39	114	**
waveform cursor	19	5.47	161	***
warping view	19	5.37	103.5	**
tolerance	18	6.44	153	****
responsiveness contrast	18	6.33	166.5	****
background colors	18	5.83	144	***
batch testing	19	6.42	190	****
6 - Developing an application for gesture sonification				
Did it work as you expected	11	5.82	64	***
video	10	5.90	42.5	**
waveform cursor	11	6.27	55	***
warping view	10	5.20	29	
tolerance	11	6.55	66	***
responsiveness contrast	11	5.73	57	**
background colors	11	5.36	30	*
batch testing	9	6.44	45	***
Did you understand the interaction between the gesture and the sound?	10	6.10	55	***
Are you able to control the system reliably?	9	5.22	26.5	**
7 - Developing an application for video scrubbing				
Did it work as you expected	10	5.80	52	**
video	10	6.40	45	***
waveform cursor	9	6.44	36	***
warping view	9	6.44	45	***
tolerance	9	6.00	42	**
responsiveness contrast	8	5.75	32.5	**
background colors	7	5.57	23	
batch testing	7	6.43	28	**
Did you understand the interaction between the gesture and the video?	10	6.00	45	***
Are you able to control the system reliably?	9	5.33	31	*

Note: The significance levels are: **** p=0.001, *** p = 0.01, ** p=0.05, * p=0.1. N is the number of participants answering a question. W is the Wilcoxon signed-rank statistic

Figure 3.7: Questionnaire results

Chapter 4

Reality-driven music interaction: using one piezo sensor to augment everyday objects

This chapter presents a tangible acoustic interface called Mogees, which gives a musical voice to physical objects by converting the vibrations created when players hit or touch them into musical sound on the fly. Users can play everyday objects as if they were musical instruments, just by using a simple piezo-transducer sensor and a mobile phone. A range of musical notes can then be triggered by physically interacting with the objects in a range of different ways. The setup of the system is immediate with the sound generated being instantaneously musical. The sound synthesis is inspired by physical rules so as to give to the instruments a sense of physicality, and its parameters can be automatically adapted so as to capture the sound of real objects. This enables users to quickly create different instruments, including hybrids applying to an object the sound created from another object. A user-study is presented to demonstrate how this reality-driven design makes the exploration of different objects and ways of playing an innate part in the interaction with the instrument.

Acknowledgements *This work is the result of collaboration with my colleague Carmine Emanuele Cella. Besides his invaluable hints and suggestions, he actively contributed to the realisation of the Mogees system and it is our intention to use the content of this chapter as a basis for a journal paper submission that will acknowledge him as second author.*

4.1 Introduction

When we think about how we interact with technology, we often still think about placing ourselves in front of a computer or a touchscreen and interacting with the virtual world inside it. A world which is realised through software and icons, and whose behaviour is governed by rules that have been designed a-priori by software developers. However, in recent years digital technology has also revealed a massive new potential for enhancing the way in which we experience and interact with the physical world around us, extending the capabilities of objects we use in our everyday lives. It is now becoming possible for example to take advantage of these everyday objects [Harrison et al., 2011a] [Xiao et al., 2013] and even of our own human bodies [Harrison et al., 2010] [Cohn et al., 2011] [Cohn et al., 2012] in order to use them as touch surfaces to control software applications in new ways and on demand whenever and wherever we need it.

We are increasingly seeing how this approach is strongly influencing the design of DMI in particular. In the last years we have witnessed a fresh blossoming of systems that embed sensors in everyday objects and use them as input devices to control music software [Rasamimanana et al., 2011] [O'Modhrain and Essl, 2004] [Cook and Pullin, 2007]. As Jacob pointed out [Jacob et al., 2008] , the greatest potential of these novel forms of TUI is to provide a *Reality-Based interaction* which takes strength through our pre-existing knowledge about the physical world around us and the physical affordances it provides to a much greater extent than was ever possible previously. However, the full extent of this potential can never be fully exploited if the digital features we super-imposed onto these objects remain disconnected from their intrinsic physical qualities and affordances (*weak* connections, using Norman's terminology [Norman, 1998]). Furthermore, if this augmentation process requires any kind of complex or long calibration in terms of the hardware and software setup, then there is a frustration for non-technical users that often limits initial uptake and sustained subsequent use of such a feature. Even for more technically skilled users, the interaction flow with the instrument will be anyway interrupted and this process of customisation of the instrument will take long time or be discouraged.

4.1.1 Sonically-augmented physical objects

This chapter presents a system that gives a musical *voice* to physical objects by converting the vibrations they make when they are hit or touched into musical sounds 'on the fly'. The physical affordances of the object are preserved because the interface becomes the object itself, whilst its sound qualities are *augmented* with enhanced musical properties. The sound is generated by connecting a vibration sensor (piezo-transducer) to a mobile phone, where the signal is used as input for an audio software engine which is designed itself to mimic the physical behaviour of real-world objects. Classic audio physical modelling works by emulating the sound produced by the interaction of a virtual *resonant body* (for example the string of an acoustic guitar) hit by a second body called the *exciter* (for example a plectrum). In Mogees, this technique is adapted so that the *exciter* is constituted by the vibration signal sensed by the piezo-transducer sensor itself. In this way, the system enhances the real-world object by instilling in it the sound properties that come through a virtual resonant body modelled in the software (*sound models*). Physical objects, from domestic tools to the trees and stones of our natural world, can be transformed so as to acquire unique, yet musical, sound properties.

4.1.2 Capturing a sound

The system allows players to quickly create new sound models using the so-called *sound capture* function, which works as follows. First, users place the sensor onto an object and hit it once; then the vibrations, which are generated by a combination of the characteristics of the object and the way it has been hit and by what, are analysed by the software which automatically sets all its parameters of the audio synthesis engine so as the virtual resonant body 'fits' the sound generated by the users' touch. In this way, any user can create, discover and improvise new sounds 'on the fly' just by selecting an object and hitting it with their hands or fingers or another object, regardless of any technical or musical skills. This function also permits a new exciting feature, the creation of *hybrid* objects: objects augmented with the sound parameters captured from another object. For example, a tree can be augmented so as to sound like a glass just by using the sound model previously captured from a glass.

4.1.3 Gestures and musical notes

The instrument attempts to give a *musical* voice to everyday objects by enhancing their natural sound properties. In order to do so, the pitch of the sound generated by the system is transposed so as to match pre-defined notes in a given musical scale. This musical scale can be set manually by users (like pentatonic, minor and so on), where the decision of which note is actually played is chosen by an algorithm and directly correlated with the gestures performed by the users. This is achieved by analysing the properties of the incoming vibrations created by that gesture, as this chapter will detail later on. The gestures of the players are therefore directly and continuously linked with the musical notes generated by the system, encouraging an engaging and sustained interaction with the physical object, which is similar to the one that musicians have with their acoustic musical instruments. By using different physical *exciters*, users can consistently generate different types of vibrations and therefore obtain different musical notes. Tapping on a table using a metal coin, a nail and a knuckle, for example, will typically generate three distinct musical notes. Unless performers wish to switch to a different musical scale or sound model during the performance, the interaction happens without the need to ever have to look at the screen.

4.1.4 Reality-driven design

In order to use our system, it *needs* to be connected to a physical object which, from then on, becomes both controller interface and sound generator. As the system works with any resonant object and only consists on a single tiny sensor, it has no fixed shape or aesthetic and its physicality coincides with the ones of the object the system is attached to. Sounds parameters are estimated directly from the vibrations of the object and immediately ready to be used. Therefore, the system is literally *plug&play* and encourages the exploration of different objects while minimising the interaction with the touchscreen of the mobile phone. The instrument does not impose any a-priori rules: different sounds and notes can be reached only through explorations of different combinations of physical resonators, exciters and performer gestures. In this way, users not only have an immediate payoff in bringing an object to musical life, but there is sustained curiosity and play in understanding the scope of what is possible for a musical performance. As this chapter aims to show, results of the user study demonstrate that the discovery, transformation and interaction with different real-world, physical objects stimulates users' creativity and curiosity and becomes a fundamental part of their interaction with the system to create sonically-augmented reality experiences.

4.1.5 Outline of the chapter

This chapter is structured as follows. First, the architecture of the system is detailed in 4.2, both in terms of interaction design and technical implementation. The algorithm for the analysis of users' input is then evaluated in 4.5 against a database, before presenting in 4.6 a user study with 17 participants from different technical and musical backgrounds. Such a study describes the effects of the design choices on the way our subjects experienced the instrument and how it impacted on their motivations for interacting with it. Finally, conclusions of this research are identified and related with the premises of this thesis.

4.2 Features of the system

The goal of Mogees is to transform the sound properties of everyday objects 'on the fly' so as to endow them with novel musical capabilities. The sound technique employed is inspired by physical rules, attempting to create a sort of physicality in the sounds it creates so as to match users expectations when interacting with physical objects. However, the goal of this system is not to *emulate* physical objects, but to *sonify* them: the sound generated by the system offers the possibility to generate musical notes coherently with users' input, in order to stimulate a deeper and more engaging interaction with the objects, similarly to the one that performers have with their musical instruments.

This work has roots in *audio physical modelling*, a family of audio synthesis techniques that attempts to digitally recreate the sound of real-world objects. This technique is generally used by specifying the list of physical parameters of a collision between two bodies: an *exciter* and a *resonant body*. In this system this technique is adapted so as a virtual resonant body is excited directly by the signal sensed by the piezo-transducer sensor. As showed in figure 4.1, users attach the sensor to the surface of a physical object and its vibrations are converted into an electrical signal which is connected to the standard audio input socket of a mobile phone and sent to the sound synthesis engine. This approach allows users to generate musical sound in *realtime* by interacting directly with the physical object the sensor is attached onto.

This section first presents a brief overview of audio physical modelling. An *audio-driven*, *physical-inspired* audio synthesis is then described, highlighting its advantages for the purposes of this thesis. The sound capture function is then presented, which allows users to quickly and easily create new sound models. Finally, the technique employed to generate musical notes accordingly to users input are detailed, which motivate this approach with the goals of this work.

4.2.1 Background: physical modelling synthesis

The idea of producing sounds by copying the physical system that emits them dates back to the 80's. The first models (Karplus-Strong [Karplus and Strong, 1983] and extended Karplus-Strong [Jaffe and Smith, 1983]) were based on the application of special filters to noise-like sources in order to create a dissipation in

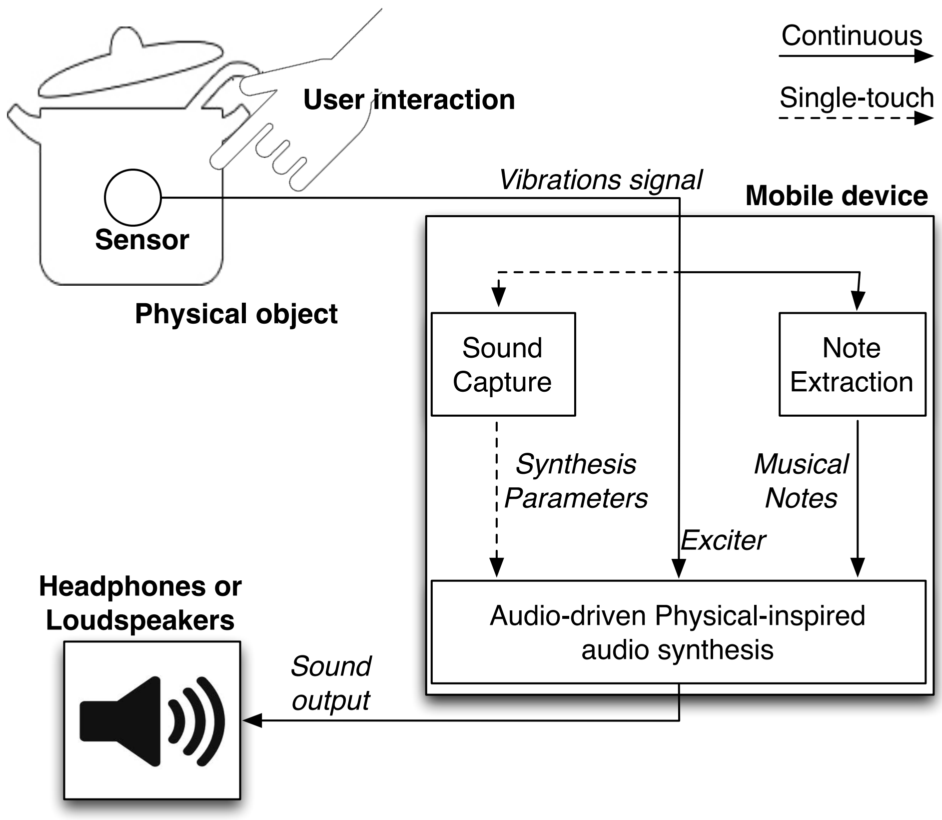

Figure 4.1: The different components of the Mogees system

the energy of high-frequency components, thus simulating the behaviour of real vibrating objects. These models are centred on the idea that a vibrating object can be represented by means of two separate interacting entities: an *exciter* (for example a plectrum or an hammer) and a *resonator* (for example the string of a guitar or a piano). Typically, the exciter injects some energy into one or more resonators that, consequentially, give some energy back to the exciter creating a non-linear system with feedback. In order to understand how this system can reproduce a physical vibrating object, some fundamental definition about mechanical vibrations is introduced.

4.2.1.1 Mechanical vibrations

A damped harmonic oscillation is a particular case of a *spring-mass* system, in which a mass is applied to a spring with a given stiffness constant. The temporal evolution of the created vibrations can be described by the following equation

$$x = e^{-\alpha t} A cos(\omega_d t + \phi) \tag{4.1}$$

where α is the *decay constant* of the system and depends on the friction, ω_d is the natural angular frequency and depends on the mass and on the stiffness of the spring and A and ϕ are the respectively the amplitude and the phase of the vibration and are determined by the initial displacement and velocity. The equation given above represents a damped vibration, also called *mode*. Despite the simplicity of the mass-spring model, complex systems can be analysed in terms of independent sets of decaying modes.

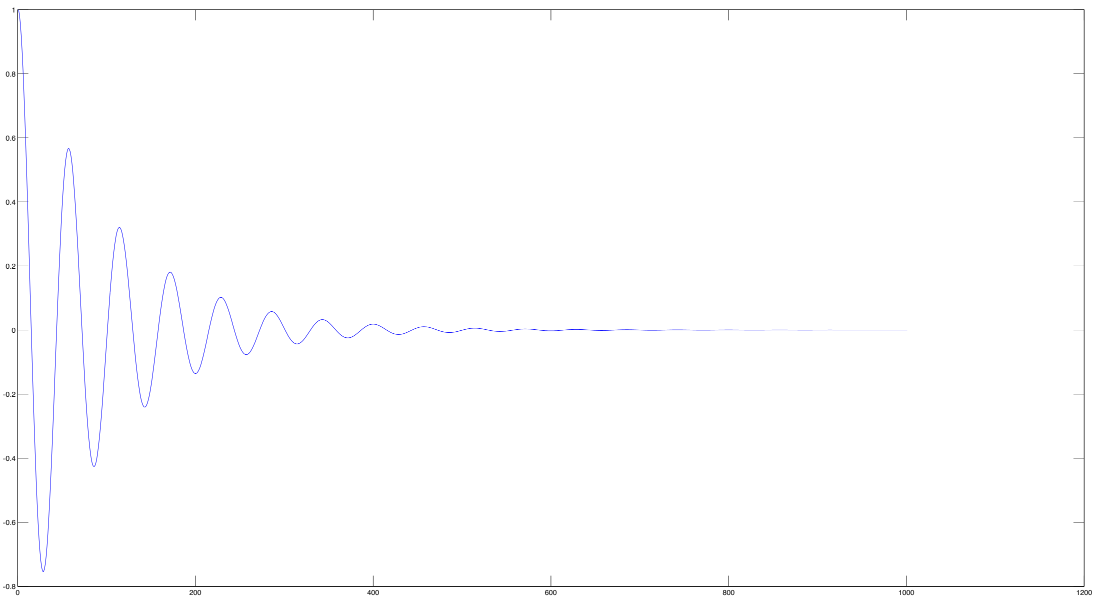

Figure 4.2: A damped vibration.

4.2.1.2 Modal synthesis and digital resonators

The complex dynamic behaviour of a vibrating object can be decomposed into contributions from a set of *modes* (damped vibrations), each of which oscillates at a single complex frequency: the generation of sounds using this approach is often called *modal synthesis*. An object that exhibits strong modes and is excited by striking or plucking is a good candidate for modal synthesis, where a resonator correspond to a single mode.

In the digital domain, the equation 4.1 can be reproduced by means of the following second-order differential equation:

$$y = x \cdot b_0 - y[z^{-1}] \cdot a_1 - y[z^{-2}] \cdot a_2 \tag{4.2}$$

where z^{-n} is the delay of n digital samples, b_0, a_1 and a_2 are called coefficients and x is an input signal; the system described by equation 4.2 is usually called a *two-poles* filter or digital *resonator* whose behavior is regulated by the value of the coefficients (figure 4.3 represents such a system).

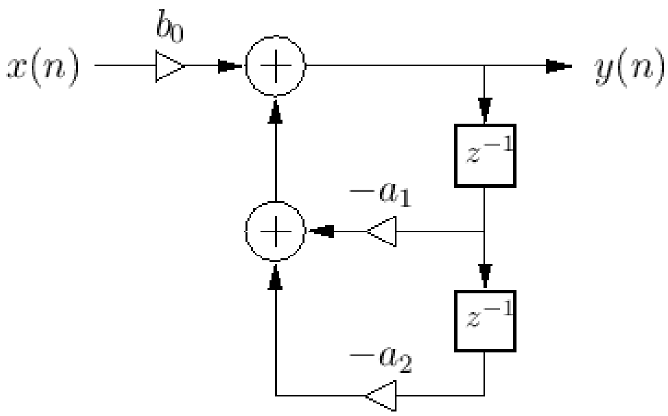

Figure 4.3: A two-poles filter.

A two-poles filter, indeed, can be designed to produce a peak at a specified frequency by setting its feedback coefficients as:

- $a_1 = -2 \cdot r \cdot cos(2 \cdot \pi \cdot f \cdot T_s)$
- $a_2 = r^2$

where r is the pole radius and T_s is the sampling period; the coefficient b_0 is consequentially computed to have a magnitude at the peak equal to 1.

A set of two-poles filters can be combined in parallel to simulate all the modes of a vibrating system; each resonator will have a different amplitude, centre frequency and rate of decay. If an exciter injects a digital impulse into the resonators, every mode will be equally excited and consequently all the amplitudes will be the same. On the other hand, if a feedback signal is added to the digital impulse, the excitation signal will exhibits a temporal smearing and a frequency equalisation; for this reason each filter will react independently to the stimulus and will assume a different amplitude, thus generating a particular *timbre*. This interaction is normally regulated by a set of weights (called *modal weights*) that are multiplied by the individual output of each resonator and that derive either from wave equations or from experimental measurements. The estimation of the parameters for the resonators is a complex matter and is usually based on experimental measurements. While the frequencies and the decay rates are based on physical properties (such as inharmonicity), the amplitudes are usually determined by the feedback interaction between the resonators and the exciter.

4.2.2 Physically-inspired, audio-driven sound synthesis

This section follows existing researches in physical-inspired sound synthesis for the simulation of the sound of real-world objects (see for example [Rocchesso and Fontana, 2003], [Cook, 1999] , [Serafin, 2004], [Testa et al., 2004] and[Avanzini et al., 2005]). The technique presented in this chapter adds important contributions to the field which are specific for the goals of this work. The realisation of this audio engine is the result of the collaboration with my colleague Carmine Emanuele Cella and represents the attempt to implement the design described in section 4.1.

4.2.2.1 Audio-driven

The Mogees sound synthesis adapts the technique described in the previous section so as to respond continuously to users interaction by analysing the stream of audio sensed by the piezo-transducer and using it directly as *exciter* for the sound engine. This approach offers several advantages. Piezo-transducers can work with a wide range of objects, sensing the vibrations propagated through the physical objects by users interaction. Thus, the signal they emit suits very well the purpose of *sonifying* users interaction with everyday objects. Unlike condenser microphones, piezo transducers are much more immune from the noise of the environment, a fundamental feature if the system needs to be used in non-silent environments such as a public space or a concert hall. Piezo-transducers are also very cheap and small and can be used with mobile phones using the standard audio input socket.

4.2.2.2 Physically-inspired

This work attempts to go beyond the pure simulation of real-world objects and transform the sound properties of physical objects whilst preserving physical plausibility.

The employed *physically-inspired* sound synthesis, indeed, permits to generate sounds that have special physical characteristics whilst not being generated by real vibrating objects.

What mainly differentiates a physical model from a physically-inspired model is the *missing* feedback interaction between the exciter and the resonators, as depicted in figure 4.4. Instead, a *parameter model* is applied to the resonators in order to create interesting evolving timbres. The system performs as follows: a continuous stimulus coming from the sensor signal is sent as input to a set of resonators, whose parameters (amplitude, frequency, decay time) derive from a specific model, as detailed in 4.2.3. These parameters are similar, in functionality, to modal weights but are not derived from physical equations. An interesting aspect of the proposed synthesis method is that it is parametric: it is therefore possible to change pitch, duration or timbre independently. Moreover, since the algorithm is tuned once but continuously fed by a time-varying signal coming from the sensor, it is possible to achieve a large variety of different sounds. This is a key aspect of the system in order to provide a satisfactory experience for the user in term of *expectations*: at every gesture performed on a object should correspond a plausible sound with some sort of physical characteristics.

Different criteria can be applied to create the *parameter model* in order to specify the frequencies, amplitudes and decay times of the resonant filters. Mogees implements an approach which allows users with no technical background to

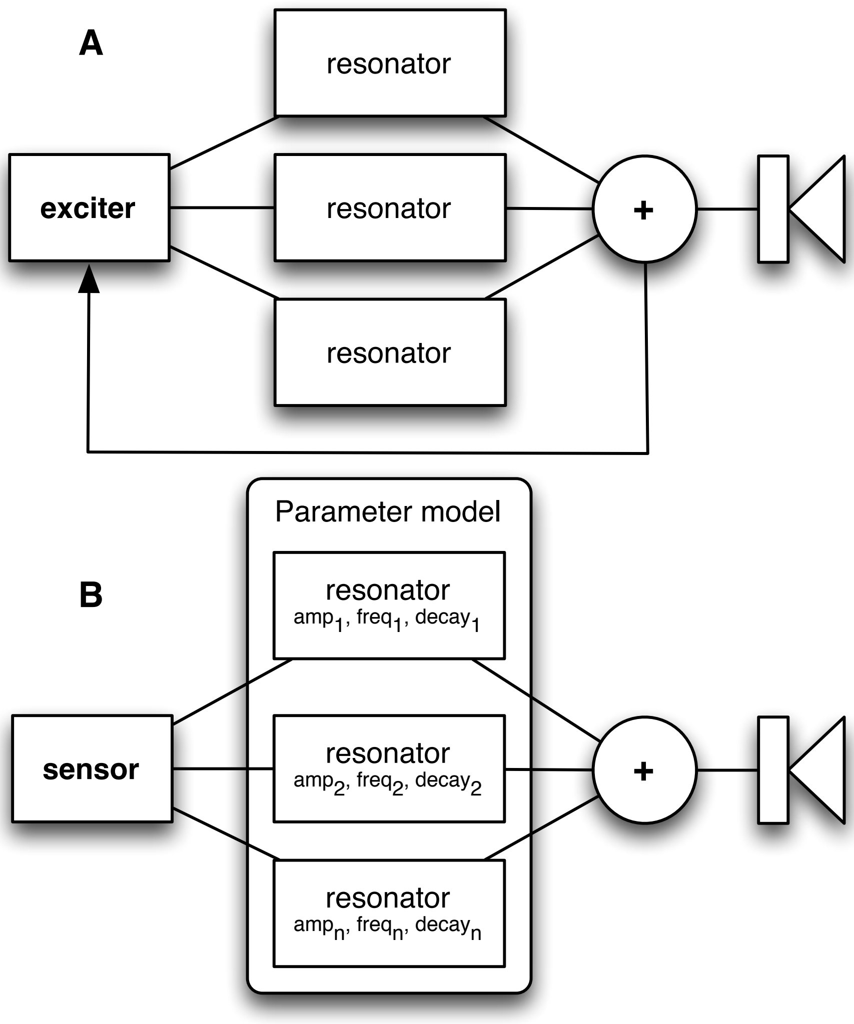

Figure 4.4: Components of classic audio physical modelling (A) and audio-driven physically-inspired audio synthesis (B).

quickly create their own sound models without having to access to any physical parameter.

4.2.3 Capturing sounds

When users want to *'play'* a new object, they can decide to play it using an existing sound preset or to create a new one which is unique of the particular object selected. This feature, named *sound capturing*, automatically sets the parameters of the audio synthesis engine based on a single example of the incoming signal.

4.2.3.1 Interaction workflow

From the users point of view, the *capture* function is presented as a single button in the graphic interface. Users place the sensor onto the objects that they want to capture and then press the button, which starts blinking, meaning it is ready to capture the next signal and build a sound preset from it. Then, as users touch the object, a correspondent onset in the incoming signal is detected and a sample of the signal is stored for processing. This process is made instantaneously and users can hear the new sound model that they created immediately from when they touched the object. At this point, they can decide whether they wish to save the sound preset with a name or to otherwise delete it and capture a new one.

This function enables a fundamental feature of the system: every object can generate a different sound, which will strongly depend on the physical properties of the object and they way the sound has been captured. Because of the immediateness of this process (plug the sensor - click the button - touch the object), this feature is designed to push users to explore as many objects as possible so as to find new sounds. This feature also enables the creation of what we could call *hybrid* objects: capturing a sound model from one object and use it to play a different one. For example, users can capture the sound of a glass and then use it when 'playing' a wooden table, creating interesting effects.

4.2.3.2 Implementation

The technique employed finds its roots in the recent techniques for Example-guided Modal Sound Synthesis proposed by Lloyd [Lloyd et al., 2011] and Ren et al. [Ren et al., 2013]. The proposed approach consists in deriving amplitudes, frequencies and decay times from the analysis of a single target sample recorded by the user, as follows:

- by using means of cepstral coefficients [Schwarz and Rodet, 1999], the *spectral envelope* of the signal is computed, which is then employed to calculate spectral peaks; the amplitudes and frequencies of these peaks are then applied to the ones of the resonators;
- a *slope analysis* is performed on the recorded sample in order to estimate the decay time of the interaction and apply it to the resonators.

The procedure is outlines in figure 4.5.

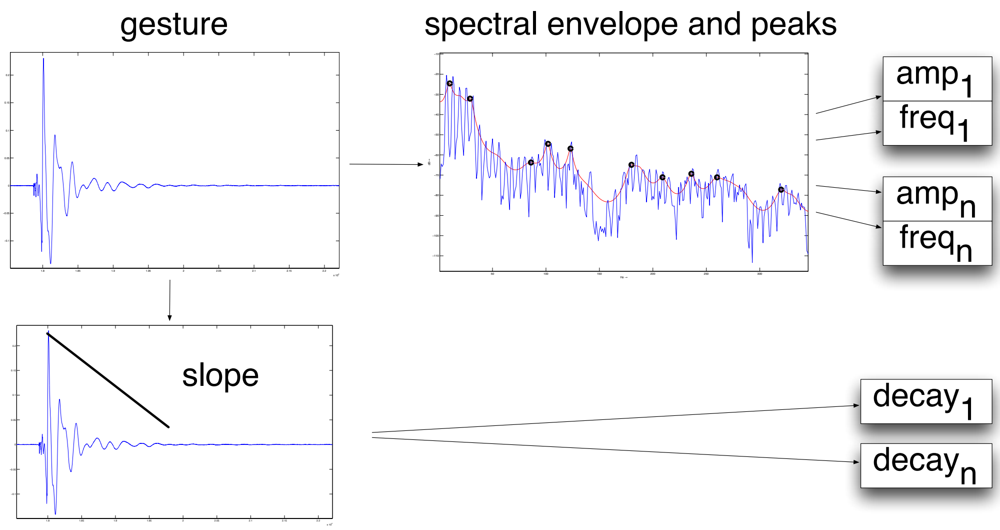

Figure 4.5: Sound 'capture' outline.

4.2.4 Playing musical notes

The algorithm described in this section has been realised with the collaboration of Carmine Emanuele Cella.

Whilst designing sounds that have a physical behaviour and that are produced by a physical interaction between the users and the physical objects is the first goal of this work, we also want the sound to be tuned to music notes coherently

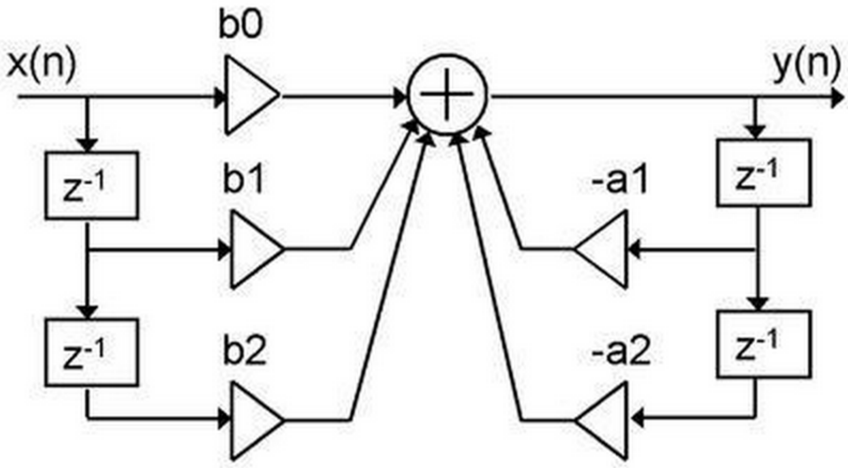

Figure 4.6: Implementation of a biquad filter

with users actions. For this reason, the outputted sound is constrained to always be tuned to a given musical scale, which can be manually pre-defined by the users.

This feature is implemented by analysing the incoming signal from the sensor using a variant of the constant-Q analysis technique, similar to the one described in [Brown, 1991] and applied for example in [Puckette et al., 1998]. In this current implementation, the signal coming from the sensor is sent to an array of biquad filters with different frequencies and constant *ratio*. Details about the implementation of a biquad filter can be seen in figure 4.6.

The frequencies of the filters are calculated so as to be more sensitive within the range of frequency where the piezo-transducer is more accurate, and in order to avoid the frequencies where the signal is more noisy (these parameters are estimated empirically and depend on the particular type of transducer employed). The frequency of the filter with highest amplitude value is then considered as the estimation of the parameter. The result is a parameter which will be higher if the incoming signal has an higher spectral content and vice versa. This approach is empirically found to be more robust than other techniques such as autocorrelation or spectral centroid for the analysis of piezo-transducer signals. As depicted in figure 4.7, this parameter is then normalised in [0,1] and then remapped to the number of notes in the musical scale selected by the user (which is a discrete number).

A realtime segmentation is performed by applying a threshold to the *slope* of the incoming audio signal. At the end of each incoming segment, the frequency parameter described above is computed. The synthesis engine contains a series of instances, one for every possible musical note, so as the frequencies of their resonators are transposed accordingly to the frequency of each note. The instance that corresponds to the selected note is fed with the incoming signal from the sensor. In order to synchronise the frequency estimation with the generated sound, a small latency is added to the sound synthesis, equal to the maximum size of a segment (1024 samples in the current implementation).

Higher notes can be triggered generating vibrations with higher frequency contents and vice versa: for example, a higher note can be triggered by hitting an object using a metal tool whilst a lower one can be triggered hitting the same object with a rubber, as depicted in figure 4.7. This produces the interesting effect of instilling harmonic properties to objects that don't have them such as the majority of the everyday objects around us. This approach establishes a continuous relationship between the actions performed by the users and the notes that are triggered, stimulating a deeper and more engaging interaction with the physical object and similar to the way musicians interact with their musical instruments. Furthermore, the number of notes that can be reached when playing a specific object directly depends on the range of frequencies that the vibrations of such objects can generate. This feature provides a further motivation for users to explore different objects and trigger different notes at each time. Unlike other analysis techniques based on supervised machine learning and matching of pre-recorded templates, however, this technique offers the advantage to avoid the imposition of pre-defined rules or gestures and also avoid any calibration, keeping the system literally *plug&play*.

A quantitative evaluation of this algorithm is described in section 4.5.

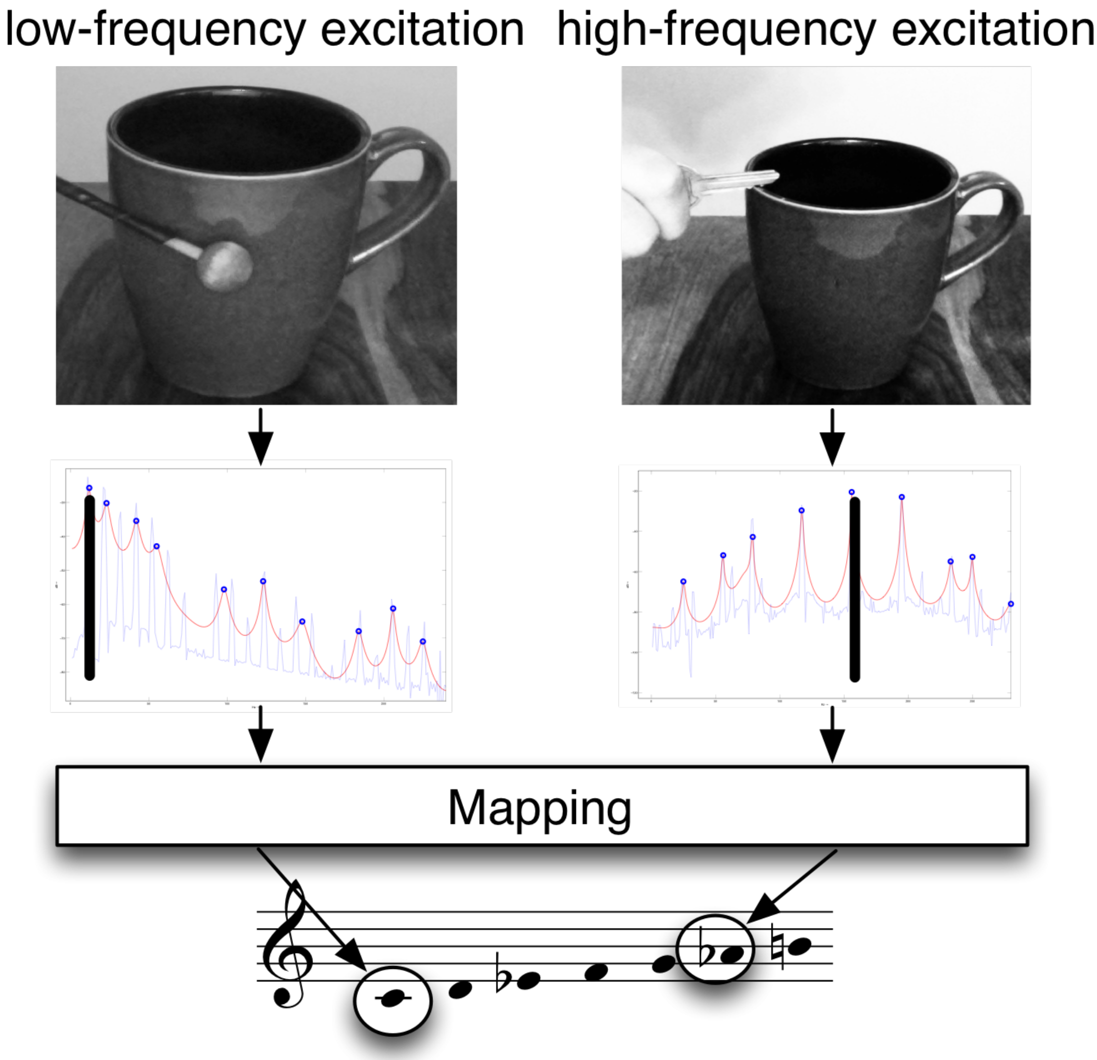

Figure 4.7: Two different types of interaction with the system that trigger two different notes in a musical scale.

4.3 Realisation

Mogees consists in a combination of hardware unit and software application. Figure 4.8 shows a typical usage of the system: the hardware unit is connected to an iPhone through a cable and attached to a resonant object (a plastic glass).

Figure 4.8: A typical usage of the system: the hardware unit is connected to an iPhone through a cable and attached to a resonant object (a plastic glass).

Algorithms implementation The algorithms described in this chapter have been first prototyped in Matlab and then implemented as a C++ audio library.

The complexity of the analysis algorithm is linear with the number of samples and directly proportional to the number of Biquad filters employed. The formula of the biquad filter is shown in figure 4.6 and consists in 5 multiplications and 4 sums, which corresponds to a total of 9 instructions. 10 biquad filters are currently employed, resulting of a complexity of 90 instructions per sample.

Likewise, the complexity of the sound synthesis is linear to the number of samples and proportional to the complexity of the two-pole filter, described in formula 4.2, which corresponds to a total of 5 instructions per sample. The number of two-pole filters employed corresponds to the number of partials of the sound created by the number of voices of the synthesiser. In a typical setup, 25 partials and 5 voices are employed. The complexity of the sound synthesis is therefore $5 * 25 * 5 = 625$.

Few things can be noticed from this result. First, the total complexity of 715 instructions per sample is fairly low compared to the processing power of modern smartphones. With a sample rate of 44100 samples per second, this requires less than 32 MIPS (million instructions per second). To give an example, the processing power is indicated to be 9210 MIPS on the iPhone 5-C and 18200 MIPS on the iPhone-S (`https://www.frc.ri.cmu.edu/~hpm/book97/ch3/processor.list.txt`). Second, both the biquad and the two-pole filters are implemented in the Apple Accelerate framework which implements a series of low-level hardware-specific optimisations. The application has been tested on the iPhone 4S, iPod Touch 5, iPad 2 and all the newer Apple devices and proved to be able to run without any audible audio-clicks or artefacts.

In terms of latency, in both the analysis and the synthesis algorithm it corresponds to the size of a single audio block. A typical setup consists of a block size of 1024 samples, corresponding to 23 milliseconds. In the user study described in this chapter no one of the participants made explicit remarks regarding this audio latency. However, this value can be decreased simply by decreasing the block size.

Application interface Figure 4.9 shows the various screenshots of the application:

1. the *sound view* allows to select the resonant model and adjust the volume of the system;

2. the *tuning view* allows to select the musical notes that are played by the system, by selecting the musical scale and the key;

3. the *capture view* invokes the capture function described in 4.2.3.

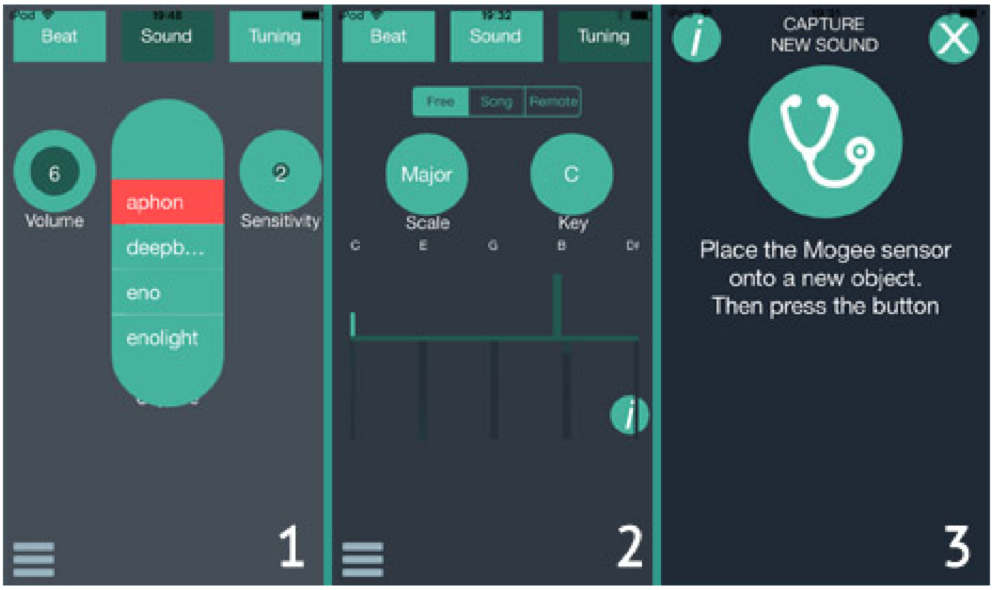

Figure 4.9: The graphical interface of the Mogees software application: the sound view (1), the tuning view (2) and the capture view (3).

Hardware unit Figure 4.10 shows the components of the hardware unit: a piezo-transducer sensor is wired to a printed circuit board (PCB) which provides connection to an audio socket. The PCB also contains two resistors of respectively 1 kOhm and 15 kOhm and a switch. These two values have been found experimentally to be the best ones in order to match respectively with the impedance of the iPhone and iPod/iPad devices. The figure also shows the aluminium case that contains the sensor and the PCB.

Figure 4.10: Components of the Mogees hardware unit, consisting in a piezo-transducer sensor, a printed circuit board and an aluminium case.

4.4 Credits of this work

Figure 4.11 summarises the credits of this work.

The concept of using piezo-sensors to sonically augment everyday objects took inspiration from other two projects developed at Ircam. First, the "MO" objects [Schnell et al., 2011] developed in the context of the Interlude project, already mentioned in the literature review of this thesis (2.3.3), that used everyday objects as interfaces for various music systems. Second, the music works with the composer Lorenzo Pagliei, with who I collaborated in several music performers for two years. These works motivated the implementation of a series of MaxMSP patches that used various combinations of audio analysis and synthesis techniques, also using the *gf* object described in chapter 3, the Ircam Modalys software for physical modelling synthesis [Eckel et al., 1995] and the MuBu toolkit for Max [Schnell et al., 2009].

These previous works contributed to the formalisation of the requirements of the Mogees system that I describe in this chapter, as well as the realisation of the first prototypes implemented as MaxMSP patch using the components mentioned above. They can be divided into sound synthesis (4.1.1 and 4.2.2), gesture analysis (sections 4.1.3, 4.5.1 and 4.5.2) and interaction design (4.1.4).

The analysis algorithm described in 4.2.4, as well as the sound synthesis technique described in 4.2.2, are the results of collaborative research together with my colleague Carmine Emanuele Cella lasted for more than two years. It is not possible to subdivide these two topic into smaller sections, nor this would help to clarify the credits of this work further. The techniques employed aim to satisfy the given requirements described and we worked together first prototoyping them in Matlab, successively implementing them in C++ and finally tested them both in MaxMSP and through the iOS app for iPhone.

The user study presented in 4.6 has been performed by myself alone. The hardware unit showed in 4.3 has been outsourced to an electric engineer by the commercial company Mogees Ltd, which I founded during my PhD thesis.

	Bruno Zamborlin	Carmine Cella
First prototype	X	
Requirements	X	
Analysis	X	X
Synthesis	X	X
Evaluation	X	
Hardware	Outsourced	

Figure 4.11: A summary of the credits of the work presented in this chapter.

4.5 Analysis evaluation

Section 4.2.4 described the audio analysis technique that is employed in the system in order to associate users different gestures to different musical notes. This section presents a quantitative evaluation of such technique in order to assess its performances against a database.

The results of the evaluation are related to the very specific requirements of the system. Therefore, sections 4.5.1 and 4.5.2 list in details respectively the goals and the constraints of our technique. The database and the methodology adopted for the evaluation are then presented in 4.5.3 and 4.5.4. Finally, the section concludes discussing the different advantages and drawbacks of this approach compared to other standard techniques (4.5.5).

4.5.1 Features

The fundamental goal of the audio analysis of Mogees is to implement a technique which can enable two distinct features:

Discrimination Different combinations of objects and exciters should produce different symbols;

Consistency The same combination of objects and exciters should produce always the same symbol.

4.5.2 Constraints

Such features, however, have to be enabled under very specific constraints. Following the general results reported in chapter 3, we list below a series of desiderata which are specific to the design of Mogees:

Unsupervised Similar to the property *user-defined gesture vocabulary* described in 3.2.2, and following the *reality-driven* paradigm highlighted in 4.1.4, we want users to be able to trigger different notes without imposing them specific a-priori rules. Therefore, for the purposes of this research the algorithm needs to work without the use of pre-recorded gesture templates . In this way, users are not forced to learn any specific behaviour in order to reach the generation of a particular symbol, such as a particular definition of *scratching* or *tapping* and so on. Instead, every behaviour they will invent during the process of customisation of the system will result in a coherent correspondence with the generated sound. Also, because the system allows users to interact with a very wide range of surfaces and thus different acoustic properties, it would have been difficult to provide a generic training for an algorithm that would work in every possible situation.

No training required A major requirement in the design of the analysis algorithm is that, for the reasons highlighted previously in this chapter, the system needed to be totally *plug&play*. Users need to be encouraged to try the system with different objects and making this process as easy and immediate as possible. Therefore, the algorithm must not require any re-calibration when applied to a different object.

Real-time Similar to the properties *Continuous Control* and *Continuous Feedback* described in 3.2.2, users' gestures need to be associated to a correspondent sound in realtime, as the gestures happen. Therefore, the algorithm needs to process incoming data in realtime with short latency.

Pitch-to-pitch mapping As the goal of the algorithm is to control the selection of a musical note, the information provided should be reflective of the pitch of the incoming signal, in order to preserve a more intuitive mapping between users gestures and produced sound. Please note that the term *pitch* is used here metaphorically and no real pitch is present in the incoming audio signal (see next point).

Robust to noise The nature of the audio signal produced by users tapping and rubbing on everyday objects is extremely noisy and lacking of any harmonic nature or pitch. It is therefore very different from the ones of a musical instrument. As no assumptions regarding any harmonic property of the incoming signal can be made, standard pitch-detection techniques such as [de Cheveigné and Kawahara, 2002] or [Chakraborty et al., 2009] are not suitable for the purposes of this work.

Efficiency The algorithm needs to run in real-time on a mobile phone. Although it has been decided to use the most powerful mobile phone commercially available on the time of this research, the iPhone 4S, its CPU limitations are still much more severe compared to the computational power of a personal computer.

This combination of required features and constraints constitute the motivation behind the choice of using the constant-Q analysis technique detailed in 4.2.4.

4.5.3 Database

In order to assess the performances of the constant-Q analysis for the purposes of this work, a database has been recorded using the Mogees hardware unit. The database consists in 11 files manually-labeled, corresponding to 11 different classes, each one containing a series of 50 instances of the same combination of material, gesture and exciter, as shown in figure 4.5.3. A class representing the combination of *glass - scratch - finger* has originally been recorded but then discarded, since the signal amplitude was too low to be correctly segmented. The files have been recorded using the hardware unit showed in figure 4.10.

4.5.4 Methodology

The evaluation methodology is structured as follows. First, the ground truth of 11 classes of 50 instances each has been created manually. Then, each instance has been segmented and analysed to create a symbol in a range of [1,11].

In order to evaluate the *discrimination* and *consistency* property listed in 4.5.1, a *clustering* evaluation has then been performed using the Rand index, which compares two different clusterings of the same database. In our case, the reference clusters are the annotate classes in the ground truth, which is compared against the results of the algorithm. In particular, the Rand index is defined as follows. Given two partitions, X and Y of the same set, the Rand index is given by:

$$Rand = \frac{a+b}{a+b+c+d}$$

where:

	Object	Gesture	Exciter
1	glass	scratch	coin
2	glass	tap	coin
3	glass	tap	finger
4	metal pot	scratch	coin
5	metal pot	scratch	finger
6	metal pot	tap	coin
7	metal pot	tap	finger
8	wooden cutting board	scratch	coin
9	wooden cutting board	scratch	finger
10	wooden cutting board	tap	coin
11	wooden cutting board	tap	finger

Figure 4.12: Left: the objects used to record the database, composed respectively of wood, metal and glass. Right: The structure of the database. Each row represents a class of the database, consisting in a set of 50 instances of different combination of object, gesture and exciter.

- a is the number of pairs of elements that are in the same set in X and in the same set in Y;
- b is the number of pairs of elements that are in different sets in X and in different sets in Y;
- c is the number of pairs of elements that are in the same set in X and in in different sets in Y;
- d is the number of pairs of elements that are in different sets in X and in the same set in Y.

The Rand index has a value between 0 and 1, with 0 indicating that the two data clusters do not agree on any pair of points and 1 indicating that the data clusters are exactly the same.

4.5.5 Results and discussion

The resulting Rand index of the evaluation of the whole database is 0.841. Being 0 the value for which the labels in the training set are totally different from the ones in the testing set and being 1 the value for which they are all the same, this result proves that the algorithm provides acceptable performances in discriminating users inputs whist keeping a good consistency.

However, the practical application of these results is to vary the gesture performed by the user without changing the material. For this purpose, figure 4.13 shows the results in the case of the three different types of material: glass, metal and wood. The rows of the tables show the gestures, whereas the columns show the label returned by the algorithm in the case of 11 classes, i.e. 11 frequency bands. As explained in 4.2.4, the algorithm can easily be adapted to output any number of symbols, simply by dividing the range of the spectrum in different numbers of bands.

The Rand indexes for these three subdivisions of the database are respectively 0.75 for the glass, 0.75 for the metal and 0.80 for the wood. From the figure, it is clear that in the case of glass, the choice of the exciter (coin or hand) clearly distinguishes between two very different notes, whereas it does not matter whether the gesture performed with the coin is a scratch or a tap. In the case of the metal, tapping with a finger clearly generates a pitch which is lower then tapping with a coin. The scratch with a coin however generates frequencies which are more spread across the high end of the spectrum, whereas the scratch with the hand produces a signal which is so low that it is not correctly segmented. In the case of the wood, scratching with the hands generates high frequencies, where all the other combinations of gesture/exciter generate frequencies that are wildly spread on the remaining part of the spectrum.

The figure shows the results of the algorithm subdividing the spectrum into 11 different frequency bands and considering the band with maximum amplitude. However, as explained in 4.2.4, this value is then normalised and linearly scaled to the number of notes available. Obviously, the lower the number of notes, the more reliable the response will be, at the cost of having a smaller set of notes available. In practical situations, the system has been used with 5 notes as a compromise. From the figure, it is easy to understand that even dividing the spectrum in 5 bands instead of 11, it would still be possible to discriminate at least two different notes per material. This number can obviously increase by applying a higher number of exciters made with different materials.

In the future, it would be interesting to compare these results with an interactive machine learning approach, similar to the one used with Gide. For example, users could record a series of examples of the gestures they intend to perform on a given object before starting to play such an object, in order to calibrate the system and maximise the results of the analysis algorithm. However, such a procedure would increase the time needed to switch from one object to another and would therefore interfere with the features described in 4.5.1.

Glass

Gesture \ Label	1	2	3	4	5	6	7	8	9	10	11
scratch - coin	0	0	0	0	0	0	0	0	0	56	4
tap - coin	0	0	0	0	0	0	0	0	0	53	0
tap- hand	47	0	0	0	0	0	0	0	0	0	0

Metal

Gesture \ Label	1	2	3	4	5	6	7	8	9	10	11
scratch - coin	0	0	0	0	0	22	1	0	3	21	47
scratch - hand	1	1	0	0	0	0	0	0	0	0	0
tap - coin	0	0	0	0	2	39	0	0	0	0	4
tap- hand	41	0	0	0	0	0	0	0	0	0	0

Wood

Gesture \ Label	1	2	3	4	5	6	7	8	9	10	11
scratch - coin	0	7	2	0	5	55	6	0	0	0	0
scratch - hand	0	0	0	0	0	0	1	0	9	0	29
tap - coin	0	0	0	0	40	0	0	0	0	3	0
tap- hand	0	9	0	0	0	30	0	0	0	0	0

Figure 4.13: Results of the evaluation of the algorithm divided by material. The rows show different gestures, whereas the columns show the labels output by the algorithm in the case of 11 classes, corresponding to 11 frequency bands.

4.6 User study

In order to assess the performances of our system from the users point of view, a user study with 17 subjects chosen from both musical and non-musical backgrounds is presented.

4.6.1 Participants

The experiment involves 17 subjects ranging in age and backgrounds. 9 of them had no previous experience with musical instruments, whilst the other 8 had a musical background, or played at least one musical instruments regularly. Amongst the musicians, 4 were familiar with digital music production software. 8 participants were females and 9 were males, and the overall age range was between 24 and 58 years old. The experiments involved one participant at the time and lasted approximately one hour. All participants took part without any incentive being offered.

4.6.2 Procedure

At the beginning of the workshop, subjects were asked to sit in front of a table where the following objects were displayed: an Apple iPhone, a part of headphones, the Mogees unit and a range of resonant objects. These objects included two kitchen pots, various silverware, a bamboo cane, metal coins, keys, a wooden vase and a remote controller for the TV. The experiments ran in a furnished living room.

Participants were asked to watch an introductory video, before assembling the system to produce their first sound. Subjects have been asked to *think aloud* while interacting with the system during the whole length of the interview. After letting them playing freely with the system for 5 minutes, subjects were asked to interact with the volume and sensitivity of the system and with the sound capture function, exploring the different objects around them. 15 minutes after, the music parameters and notes visualisation in the graphic interface are showed to the participants and the note generation is discussed, watching the difference of their approach with the system compared to the previous phase. Participants are then asked to fill a System Usability Scale (SUS) questionnaire [Brooke, 1996].

Finally, the semi-structured interview follows, of which the list of questions is reported below:

- What do you think of what you've looked at today?
- Do you find that the sound corresponds to your actions or not?
- What do you think of the sounds that you created? Do you think that they correspond to the physical object or not?
- Can you predict how the system would behave with different surfaces/objects or are you surprised by the sounds that you hear?
- Are you encouraged to explore different types of objects? If so, what are the main motivations?
- Do you think that the system could be used for different applications other than musical performance?

4.6.3 Results

The results of the user study are now presented, based on the information collected through the semi-structured interviews and the System Usability Scale (SUS) questionnaire, whose results are shown in table 4.1.

The average score for musicians is 81/100, for non-musicians 74/100. Some notable outcomes of the SUS are that subjects were sure that users would understand how to use the system quickly (4.31), they did not find the system unnecessarily complex (1.69), they would not need the support of a technical person (1.25) and they think the functions of the system were well integrated (4.44). However, non-musicians gave a significantly lower rank to the point 'I think that I would like to use this system frequently' than musicians (calculated with a two sample Wilcoxon rank sum test $W = 52.5$, p-value $= 0.024$) and four subjects scored 4 / 5 in the question *'10-I needed to learn a lot of things before I could get going with the system'*. These results are discussed below and combined with the ones gathered in the semi-structured interviews, grouping them in three main subjects.

Question	Musicians		Non-Musicians		All	
1	3.71		2.33	*	2.94	
2	1.43	**	1.89	**	1.69	***
3	3.86	**	3.11		3.44	*
4	1.00	**	1.44	***	1.25	****
5	4.29	**	4.56	**	4.44	****
6	1.57	**	1.22	***	1.38	****
7	4.14	*	4.44	**	4.31	***
8	1.14	**	1.22	***	1.19	****
9	1.14	*	1.22	*	1.19	***
10	2.00	*	2.56		2.31	**

Table 4.1: Results from the System Usability Scale questionnaire. All values means of the participans responses in the range 1 to 5. The first column lists the question number, corresponding to the following: 1 - I think that I would like to use this system frequently 2 - I found the system unnecessarily complex 3 - I thought the system was easy to use 4 - I think that I would need the support of a technical person to be able to use this system 5 - I found the various functions in this system were well integrated 6 - I thought there was too much inconsistency in this system 7 - I would imagine that most people would learn to use this system very quickly 8 - I found the system very cumbersome to use 9 - I felt very confident in using the system 10 - I needed to learn a lot of things before I could get going with the system. Significance levels were calculated using a one sample Wilcoxon signed-rank, measuring the difference between the sample responses and the mid point of the scale (3) and are given as: **** $p < 0.001$, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Full data are showed in figure 4.6.3.

4.6.3.1 What our subjects think about the sounds they created?

At the beginning of the interviews, the subjects have been asked to freely explore the system with the different objects available, so as to evaluate how the sounds produced by the system were perceived. Generally, all subjects agreed in feeling a sort of *physicality* of the sounds they created and in the sensitivity of the system, a direct and predictable correlation between the physical objects and the sound models they *captured* from them, a shared interest in creating *hybrid* objects and a (positive) sense of artificiality in the note generation.

Almost everyone started exploring the system by sticking the sensor onto the table and tapping as if on a drum or a keyboard. The question *'Do you find that the sound corresponds to your actions or not?'* collected all positive answers, and several subjects used words such as *authentic*, *coherent* and *real* to describe such a correspondence (a non-musician, at the end of the interview, even asked us to explain *'where the trick was'* wondering whether the sound she heard *was real or not*). Overall, musicians were the first to discover that the system was responding also to gestures other than tapping, such as by continuously scratching onto the surface of the object. With the exception of 5 subjects, who had already seen the demo video and live performances of the first author using the system before the interview, the remaining 12 expressed signs of pleasurable surprise when they heard the first sounds being produced by their touch. 6 subjects, all non-musicians, needed to be told that they could interact in ways other than tapping. They all displayed positive surprise when they realised that this was possible. Two musicians were worried about this high sensitivity when they discovered that audio was being generated accidentally by them walking on the floor whilst the sensor was plugged onto the table, arguing how this could be a potential issue in live performances. However, they have been satisfied when I explained that the sensitivity of the system was controllable through the touchscreen interface.

Interestingly, all 9 non-musicians spontaneously used at least one word related to physical properties when describing the sounds created with the *capture* function. For example, words like *light*, *cold* and *thin* have been used to describe the sound captured from a glass, *creaking* and *heavy* in order to describe the sound captured from an old wooden table, and *feel* to describe the interaction.

Our subjects were also asked whether they thought that they could predict how the system would behave with different objects, or if they were surprised by the sounds that they heard. With the exception of 2 subjects, who were unsure, the other 15 claimed that they could probably predict what a new object would sound like. 6 of them also argued that this skill would probably increase with a prolonged use of the system. 3 musicians, which were used to audio production software, asked if it was possible to change the sound using parameters in the touchscreen, arguing how this would lead to a more precise control of the sound (for example, a laptop performer claimed *'I'd prefer a wider range of sounds'*). However, he agreed when pointed out that the audio could be easily routed to another software for post audio effects. All the other participants, however, were excited about the idea of creating sounds by exploring different objects.

The possibility of creating *hybrid* objects, i.e. capturing the sound of an object and then using it when playing on a different object, has been spontaneously explored by every subject. Amongst them, 7 subjects (2 non-musicians and 5 musicians) explicitly recognised it as a compelling feature.

The pitch transposition and note generation has generally been perceived as a sort of gaming artefact, an audio effect applied to a more realistic sound generated underneath. For example one subject, also researcher in audio technology, referred to a sense of *artificiality* in the generation of the musical notes by saying: *'Yes the sounds correspond a lot to the objects, but they are pitched'*. As discussed in the next paragraph, such an artefact has positively resulted in a sort of *task* for the interaction of both musicians and non-musicians, driving curiosity and questions about the way they were produced.

4.6.3.2 Playing the notes

Although the system is designed so as to produce sounds which are always tuned to a musical scale, triggering different musical notes reliably proved to be a challenging task to achieve. Musicians were generally interested in reproducing specific melodies, but they did not succeed immediately. Further explanation and examples, however, generated excitement toward the system, a more intense interaction and confidence in improving their performances in the longer term, as well as willingness to explore different objects and techniques. On the contrary, non-musicians showed more curiosity in exploring different objects and discover new sounds rather than reproducing specific melodies. They also spontaneously suggested applications for the system in music pedagogy.

Non-musicians were globally more intrigued by the other features of the system and did not pay too much attention to the reproduction of series of notes. For example, 3 non-musicians explicitly stated that they see the system, rather than a pure musical instrument, as an exploratory tool to discover new sounds *hidden* in the physical objects (one of them referred to a stethoscope as a metaphor for the system). 5 subjects described the possibility of controlling musical notes as *challenging* or *entertaining*, also claiming that the correspondent visualisation in the touchscreen was providing a clearer *task* for their actions and the impression of *learning something useful*. 5 of them spontaneously discussed potential applications in music pedagogy, specially for children. 3 of them motivated this answer with the simplicity of the system (*it's so easy that I'm sure my son would use it*) and its portability (*you can exercise while you're waiting the bus*), whilst 2 of them mentioned a sense of immediate reward, probably also linked with the decision of triggering notes harmonically correlated (*it makes me feeling like a musician*). This idea shared by several non-musicians about the system being useful for music pedagogy is linked with their low rank in the question *'1-I think I would use the system frequently'* (2.3), implicitly meaning that they were not interested in improving their musical skills. This might been due to the fact that the main output of the system is a freeform musical one, with no *gamification* having been introduced into the app. Interestingly, this view differs from the one shared by other 3 non-musicians which described the system as an exploratory tool, and rated this question between 3 and 4.

Unsurprisingly, musicians were instead much more concerned in reproducing series of notes reliably to their gestures. When I showed them the notes visualisation in the touchscreen, which visualises the notes that users are triggering over time as well as the possibility of changing the musical scale, this led to a general excitement and a more intense interaction, and they saw this visualisation as a tool to improve their performances. 2 of them noticed that, by deciding what range of notes they could play, they could easily perform together with other acoustic instruments. 4 of them expressed frustration when trying to reach the same sequence of notes using bare hands for few times without success. When a short demonstration has been provided pointing out the possibility of triggering different notes by the use of different exciters (such as metal coin, a knife and different part of the hand), however, they became more engaged and intrigued by such a possibility.

The challenging aspect of controlling the notes reliably, as opposed to its simplicity in using different objects and creating new sounds, conveyed to a desire for long term use of the system by musicians. With the exception of 1 subject, all musicians believed that their skills in controlling the system could improve substantially over time, which was probably one of the motivations behind the high rank to the question *'I think that I would like to use this system frequently'* (3.7 / 5). One subject, who regularly plays percussion, said: *I'm sure I could do a lot with this app using different objects and materials. It lets me make discoveries (...) Other software are more enclosed, here the results really depend on what you do'*. Similarly, another subject said *'There is an element of practice, you get better by learning how it reacts to different touches'*.

The interaction with the touchscreen have been predominant during the first 10 minutes of the interaction for 3 musicians and 4 non-musicians, mainly to explore the sensitivity and volume knobs as well as to change the sound presets. 3 users expected the system to generate sound by their interaction with the touchscreen alone. However, after the first phase of free experimentation, the interaction with the touchscreen became increasingly sporadic. With the exclusion of the capture button, this second phase showed that our subjects hardly changed any other parameter at all. 5 musicians specifically claimed that the graphic interface, specially the tuning view, was useful to learn how the system works but they would probably stop watching it when performing in the longer term.

4.6.3.3 Why exploring?

The final part of our results section focuses on a more general discussion about system usage, discussing the strategies adopted by our subjects, their interpretations of the system and their speculations about longer term uses. The findings suggest that they considered *exploration* of different physical objects and environments as an inner activity with the system, with the discovery of novel sounds and original game-plays as the main motivations behind it.

When our subjects have been asked to try the system with different objects, only 2 of them limited the explorations to the objects that they found on the table. The other 15 spontaneously explored furnitures present in the whole room, whilst 3 of them also walked out of the room. The most commonly explored objets were a glass window, an old wooden wardrobe and a chest. Two musicians were also interested in exploring different parts of their own body, both finding that the bones produced the most interesting results (probably because the skin tends to attenuate the vibrations). The results of this study support [Cook and Pullin, 2007] in observing enthusiasm and a great variety of objects that users would like to play. 11 subjects mentioned that they would like to try the system under conditions that were not available in the room in which the experiment took place and spontaneously described different objects and environments. Amongst these, there was a clear predominance of natural environments (rocks, stones, trees and even a volcano) and objects present in everyday activities (bus shelter, train floor and windows, car indoor furnitures). These descriptions shared a general interpretation of such *explorations*, by both musicians and non-musicians, as an innate component in the usage of the system.

When subjects were asked to discuss the main *motivations* that would possibly drive them to explore different objects, the most common answer was the desire to create new fresh sounds (*'The beauty of this system relies on one playing places that you wouldn't reach otherwise'*). These answers were often linked with *curiosity* about how specific objects or elements would sound like (*'I'd be curious to know how a volcano would sound like'*). Musicians favoured artistic motivations related to aesthetic and originality (*playing the trees in a forest*), as well as adding elements of variation and site-specificity in their concerts by playing objects linked with the location where the concerts happen. 3 non-musicians mentioned the size of the object (*'I'd like to try to play a car because it is big and gives me many possibilities'*) whilst other 3 suggested the possibility to use the system to generate music as soundtracks of their journeys, such as putting the sensor in their pockets and sonify a walk or putting it on the floor of a train (*'I wish I had this so to make sounds from the noise of the engine of the train I took this morning, it would have made my journey more fun'*).

Amongst the most recurrent application scenarios for the system suggested by our subjects, there has been a clear predominance for music education, specially for children, mentioned by 5 musicians and 4 non-musicians. Other common topics have been game applications for people with disabilities, interactive systems for dancers and video game controllers.

4.7 Conclusions

This chapter described a new system that implements an audio-driven, physically-inspired sound synthesis, showing how this technique can allow users to create musical sounds from everyday objects and activate engaging musical interactions.

Instead of enabling the users to explore and define different gestures, Mogees' users customise the interaction with the system by exploring different physical objects. The gestures they employ are invented by the users and limited by the only fact that they have to produce sensible vibrations composed by different frequencies. The *reality-driven* design proposed is designed to provide unique sounds based on the physical object it is attached to and the gestures its players perform. In both Gide and Mogees, the usage of the system requires an initial customisation of the modes of interaction: in Gide, this is done by recording gestures in the system; in Mogees by selecting a physical object and the gestures to play with it.

This study showed that the possibility of creating new sounds from the acoustic nature of physical objects proved to be compelling for both musicians and non-musicians, and the sound generated by the system to be predictable and coherent with their gestures, supporting the *visibility* property described by Alan Dix. Visibility proved to enable participants to focus their attention almost entirely to the interaction with the physical object, often forgetting the screen of the mobile phone until they were requested to change the setting, which is considered to be a compelling feature for a digital musical instrument:

> *When we find a performer that does not care about the computer screen display, when we see someone on stage capable of lovely caressing and of violently striking the instrument without any fear, chances are we are facing a memorable performance.* [Jordà, 2004a]

Users showed to be highly motivated in exploring the system with different objects providing unique and personal motivations, ranging from objects aesthetic and size, to the desire to create unique instruments, to simple curiosity. Exploration has been considered an inner part of the interaction workflow with the instrument and the sound capture function proved to enable users - irrespective of their musical or technical skills - to quickly create sound models which are realistic and coherent, as well as to create compelling *hybrid* objects.

Although the system proved to be immediate to setup and pleasant to play from the very first minutes of usage, the level of note control is different than the one of other musical instruments and might require specific training. Once musicians assisted to a demonstration of myself succeeding triggering different notes, however, they acquired a stronger desire to practice with the system further so as to reach a better control of the instrument, suggesting the potential scope for a sustained engagement and opportunity for virtuosity.

This *need for guidance* is an important result of the research. Whereas Gide provides precise visuals and acoustic feedbacks to the users guiding them in the process of defining and performing their own gestures, in Mogees such guidance is absent. Users are free to explore with an endless range of possibilities but no hint is given beforehand about what such possibilities are. These results supports Carroll's guidelines [Carroll, 2004]: there is a risk of usability in malleable technologies as their flexibility can cause a lack of constraints and absence of a clear guidance on the use of the technology. In the study, however, a simple practical demonstration succeeded in illustrating the features of the system without limiting the creativity in its usage.

The positive results of the user study described in this chapter in terms of usability and stimulus for creativity motivated the desire for a second study, presented in the next chapter, for which such a balance between flexibility and clear guidance reveals in all its potential: the case of music education with young children.

USERS (MUSICIANS)	1	1	3	4	5	6	7	8	9	10
A	3	1	3	1	4	2	2	2	4	2
B	5	1	4	1	5	1	5	1	4	1
C	4	2	4	1	4	1	4	1	3	1
D	3	1	5	1	5	1	5	1	5	2
E	2	2	3	1	3	2	5	1	3	4
F	4	1	4	1	5	1	4	1	5	3
G	5	2	4	1	4	3	4	1	4	1
USERS (NON MUSICIANS)										
H	3	1	3	1	5	1	5	1	4	1
I	2	1	4	1	5	1	3	1	3	4
J	2	3	2	2	5	1	5	1	3	3
K	3	2	3	1	4	1	5	1	3	2
L	2	1	5	1	5	1	5	1	4	4
M	4	2	2	1	3	2	5	2	5	2
N	2	3	3	2	4	2	4	2	3	2
O	1	3	3	2	5	1	5	1	3	4
P	2	1	3	2	5	1	3	1	4	1
MEAN (MUSICIANS)	3.71	1.42	3.85	1	4.28	1.57	4.14	1.14	4	2
MEAN (NON MUSICIANS)	2.33	1.88	3.11	1.44	4.55	1.22	4.44	1.22	3.55	2.55
MEAN	2.93	1.68	3.43	1.25	4.43	1.37	4.31	1.18	3.75	2.31

Figure 4.14: Full results of the questionnaire presented in 4.6.3.

Using Carroll's terminology, a technology-as-designed (as provided by a designer) is a different matter to the technology-in-use, the one embedded into the lives of its users and appropriated into their daily tasks and routines. Too often, in the field of new interfaces for musical expressions, the research focuses on the former and observations about the latter are limited to short user studies in controlled environments. The goals of this work is to observe and enhance such a variation between these two states of a music technology and observe how decisions made at design time affected the usage of the technology by its players in real world situations.

During the time of this work, I spent three years producing and distributing instances of the Mogees instrument to different artists, from percussionists to laptop musicians, from Kathak and tip tap dancers to flamenco guitar players, from music educators to Cabaret showmen. Each time I tried to engage with the users, providing them technical support and watching how the system was appropriated and used in a such heterogeneous spectrum of scenarios.

Although most of these conversations happened informally and spontaneously, two particularly important experiences, radically different one from another, have been framed into research studies and build the second part of this thesis. Chapter 5 presents a study in a primary school, watching the Mogees system being employed in the field of music education for pupils ranging between 7 and 9 years old. On the contrary, chapter 6 describes the adoption of the instrument by a seasoned electronic music producer, during an 18 months long collaboration including both studio work and live performances.

Chapter 5

Evaluating Mogees for music education in primary schools

5.1 Introduction

During the last year of my research, I collaborated with the european project PRAISE (Practice and peRformance Analysis Inspiring Social Education) [pra] to evaluate the potential of using the Mogees system in primary schools. During this period, I collaborated with the music educationalist Nancy Evans, as well as with the Goldsmiths researcher Harry Brenton, to assess the potential of the Mogees system to facilitate access to music-making and music learning for young people, particularly those where traditional instruments are not accessible.

This provided an ideal context to evaluate the usage of the Mogees system in a very defined and established practice: the one of music pedagogy for young people. This section analyses the results of this study, as well as reporting the impressions and thoughts expressed by the educationalist herself in relation with the technology. The goals of this section are thus twofold: firstly, to assess how children interacted with the system and what they learned from it, evaluating the potential of Mogees in music education for 7-9 years old pupils; secondly, to describe the usage of the system from the point of view of a real user, the music educationalist, reporting the way she used the system and the impressions she expressed about it. This provides interesting insights about how the system has been adapted and customised by the pupils and the educationalist herself in a practical context of use.

Credits of this work This chapter describes and analyses the results of a study in a primary school in Birmingham with 9 pairs of pupils between 7 and 9 years old. The methodology of the study has been originally designed by myself and then reviewed by Nancy Evans and Harry Brenton. The study has been conducted by Nancy Evans and myself on October 24, 2013 and by Nancy Evans alone on October 25, 2013, with the support of the PRAISE project [pra]. The data collected during the study have been analysed by Nancy Evans and myself and formed the basis for an informal report written by her few days after the study and included in this thesis as appendix B. Such a report, as well as various meetings and phone calls, extended the results of the study with the personal vision of the educationalist herself regarding a potential employment of the Mogees system in primary schools.

Chapter outline This chapter is structured as follows. It first presents the context of the study and the participants that have been involved. It will then lists the aims and the objectives of the study, as well as detailing the methodology employed. Then, the observations collected by the educationalist during the study will be reported and integrated with her personal assessments about the strength and weaknesses of using the system in this particular context. Finally, the section drives some conclusions in relation with the goals of this work.

5.1.1 Context and participants

9 pairs of children between 7 and 9 years old from Highters Heath Community Primary School took part in the study. All the children access computers/iPhones/iPads etc. regularly at home and at school. Three children used them every day, three most days and twelve children sometimes. The most popular use of technology is to play games (18), then looking things up (14), then watching watch films/dvds etc. (13) and finally Wii (12). Eight of the children said they had used computers/iPads/iPhones to create music, including Singing Mario, Bruno Mars Lazy, Mini Piano and Make Music 95. Others described unnamed karaoke programmes or described recording themselves.

5.1.2 Aims and objectives of the study

The aims of the study are twofolds and take into account two different users: the students and the teacher.

First, we want to build a child's view of the Mogees system and assess its learning potential in music education for 7-9 years old children. In order to do so, we observed and recorded the different strategies and activities that children adopted when they interacted with the system, their level of engagement and their improvements during the class. The study is hands-on and allows for recording any technical and practical difficulty of using the system in a real world primary school during the schedule of a normal school day. This served to assess the level of support required from the school and the level of preparation required for the teacher in order to embed the system in the music teaching.

However, there is also another important outcome of this study: to build a view of the system from the eyes of the music teacher. At the end of the study, the teacher has been asked to make an initial assessment about the strengths and weaknesses of using Mogees with young children, generalising this experience to a broader context.

5.2 Methodology

The aims and objectives listed in the previous subsection provided a guideline for the protocol of the study described in this subsection.

5.2.1 Technical setup

The study has been conducted in a small classroom, set up as in figure 5.1. We made available a series of physical objects selected to be used as *exciters*. These included a coin, paper clip, screw, paint brush, green plastic leaf, chopstick, whisk, nail, ping pong ball, pickle stick, large metal disc. In the other side of the room, we sticked to a desk a Mogees unit, which is composed by the hardware sensor and an Apple iPod 5.

5.2.2 Participants

The children took part in the experiment in pairs. In total there were 11 pairs. 2 pairs have been used for a pilot, whereas for the other 9 pairs the definitive protocol has been employed. Each session lasted between 30 and 40 minutes.

5.2.3 Tasks

Introduction On arrival, the children were shown the different bits of the hardware: iPod, speakers and contact microphone. It was explained to the children that Mogees was a new app that turns the table and every object into a musical instrument and that they would be able to make different sounds using different ways of playing with their hands and with the different *exciters* available.

Free exploration The children were given about five minutes to freely explore Mogees with no intervention from the adult(s). Any dialogue they had was noted and observations of what they did with the exciters/exciters taken: (a) what kind of gestures did they make? (b) were there particular exciters that they find more satisfying than others? At the end of the five minutes the children were asked individually to play some of the music they had created.

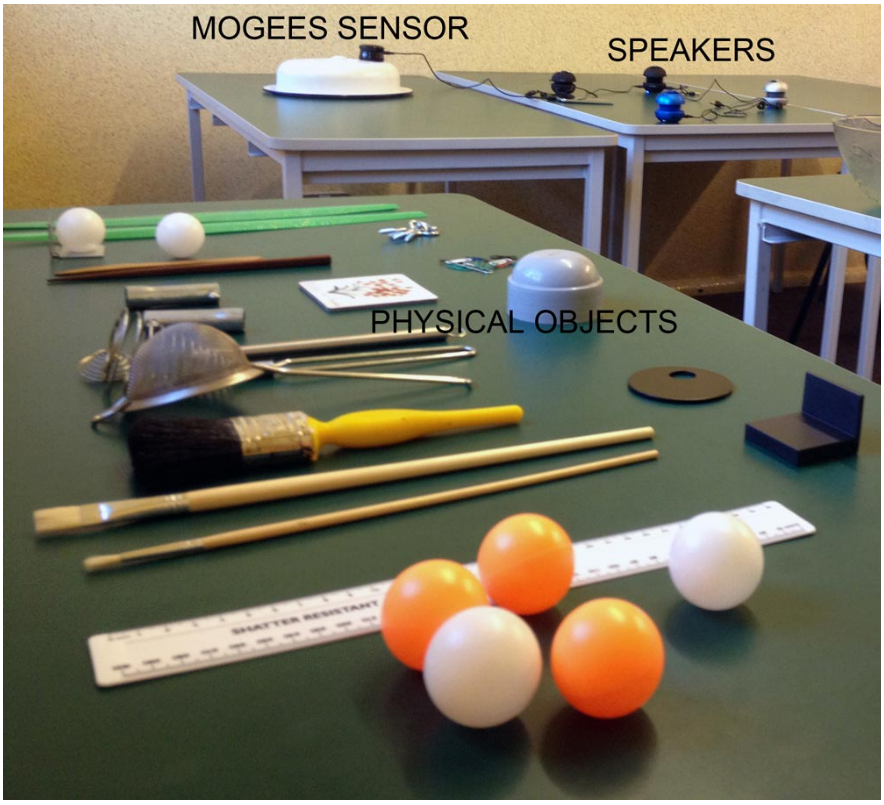

Figure 5.1: The setup of the user study in the Highters Heath Community Primary School in Birmingham. Several physical objects were available for the children to use with the Mogees system.

Making rhythms The children were then asked to play a short repeating rhythm on Mogees. To help with this, children were asked to practice clapping simple rhythm suggested by the adults before playing it using Mogees. It was noted down whether the children were able to reliably repeat the pattern.

Expert demonstration At this point there was a short Mogees demonstration from an adult (either Nancy or myself) that included the use of hands.

Free exploration of the interface The children were shown how to change the sound on the iPod interface, allowed to change the sound and continue their explorations. After a while they were asked to plan making some music together.

Composing a piece as a duo The children as a pair were then asked to create some music using their favourite sound and exciters.

Testing the usability of the graphical interface Next, children were asked to change the sound preset, volume and music scale of Mogees to specific values. The children were told they could ask for help if they stuck. It was noted how often they asked for help and the questions they asked.

Using the capture function It was explained to the children that so far they had been using sounds already in the app and that now they would be able to *capture* their own sound from an object. Each child was then asked to choose an object and use it to capture a sound. It was explained that depending on how they played the object and which exciter they used, they would capture a different sound. The children were shown how to use the graphical interface to accomplish this task and then ask to do it by themselves. Finally, the children were asked individually to make some music playing the table with their new captured sound.

Questionnaire Before the beginning of the experiment, the children answered to a background questionnaire regarding their pre-existing music skills and experience with apps (see B for the list of questions and answers). At the end of the experiment, a second questionnaire about their experience with Mogees has been presented.

Every session has been video recorded. The following sections of each video have been analysed by the teacher: the initial exploratory stage, the individual improvisations, the music created by the pupils after the teacher's or my demonstrations, the capturing sounds and individual improvisations.

5.3 Discussions

The discussions presented in this section are the result of the analysis of the report provided by the teacher, included in appendix B. The relevant observations have been divided into five main topics that are relevant for the purposes of this thesis.

5.3.1 Immediate reward, need for practice, room for virtuosity

Mogees proved to be simple to understand and to use from the very beginning. The terminology used by the children to describe their understanding of the system suggests that the concept of applying the sensor to objects and making music out of it was simple and easily understandable: '*It takes sounds and makes an even better sound from it*', '*Makes music and sounds using the microphone*' and '*Takes the sounds of stuff*' were some of the answers given by the children. The teacher also reported that no children had problems in using the software interface. The main difficulties reported has been about using exciters which did not trigger hearable sounds, such as soft brushes. However, '*All children said that they thought the difficulties would improve with practice and when they had tested them*'.

All the children showed to be curious about the new instrument and willing not only to learn to create novel sounds with it but also to suggest novel way to use it. Suggestions have been rich and creatives, such as creating sounds of animals, music instruments and spoken words, using the music produces as a ringtone and sharing them with friends.

As noticed by the teacher, '*Mogees allows for complex interactions and expressing creativity without difficult learning barriers such as, for example, to bow a violin or get a correct embouchure on a trumpet*'. The fact that the sound produced is constrained to a musical scale also encourages to play further and to develop satisfaction. However, like any musical instrument, the musical result improves within time and has its own learning trajectory. Trying out new techniques, exploring sound quality, listening to oneself play, repeating over and over, building up new musical ideas and revisiting music made before: all these activities require time. This point has been noticed by the teacher during the study, as a very useful reminder that Mogees, like any conventional instrument, despite producing what can seem like immediate success/satisfying sounds, needs practice.

The teacher reported that *within this short session, Mogees did sustain the attention of all the children.* However it was not possible, within this very short time, to ascertain whether this attention would continue over a number of sessions.

5.3.2 A rule-free environment

The children often used the exciters that were available in a way which was suggested by the common usage of such objects. For example, scratching the table with a knife, brushing it with a paint brush or a sponge, bouncing a ping pong ball, tapping with a stick.

Whereas traditional acoustic instruments have been designed to afford a specific set of actions from their users in order to produce a range of sounds, Mogees offers a bigger and more diverse set of gestures within one instrument, and such gestures can be suggested by the objects themselves. As long as a gesture produces a given sound through the object the sensor is attached to, it can be incorporated in the practice with the instrument. From the way the instrument is presented, i.e. a sensor that waits to be attached somewhere, the concept of physical affordances is open to users' (and childrens') imagination in their decision about what they want to play and how.

Of course traditional instruments can also be modified and played with alternative techniques and using different varieties of objects (see the prepared piano as the most classic example). However, this behaviour in Mogees is not a possibility but a requirement, and users are pushed to select the combination of resonator/exciters in order to play. The impact of this decision in the generated sound is essential in producing a variety of different notes.

In Mogees the concept of affordances is initially more open and as yet less explored compared to traditional instruments, as the system can be used with a vast range of objects and materials. This could be seen as a potential problem in an educational context, due to the freedom that the instrument allows and a lack of clear guidelines. However, the study suggested that this freedom has been compensated with the pre-existing knowledge that the children had regarding the objects they had at their disposal and the way they can interact with them. This association acted as a trigger for the imagination of the pupils and motivated them to develop ideas and a variety of possible scenarios of play.

5.3.3 Finding the notes

As reported by the teacher, "one of the most notable differences between Mogees and traditional pitched instruments like a piano or a guitar is that in the former the notes are not spatialised. In order to trigger a different note, in Mogees it is required to change the type of the gesture rather than performing the same gesture in different locations". Therefore, teaching how to trigger different notes can be considerably different than traditional music instruments and can vary from object to object and from player to player, making the process of customisation more evident. In this educational context, such a process does not concern only the act of triggering different notes reliably with a given object, but rather how to *discover* notes when a novel object is selected, and understanding the reasons behind it.

The children adopted a trial and error approach to find different notes, by experimenting with different exciters and different gestures. This approach is interestingly very similar to the more expert approach adopted by myself when exploring new objects, as detailed in the next chapter (6.3.4). This necessity acts most of the time as a great motivation for the pupils to find and invent new solutions to play. Every kid has been willing to explore different objects and most of them were passionate in the solutions they found. It happens few times that this approach caused a sense of frustration from the pupils and the teacher needed to intervened suggesting to play in different ways in order to achieve different notes ('*I can't find the sound I'm looking for*').

Some children also tried to play a song that they already knew, reminding us that imitation is a great motivation to learn a music instrument. This study suggests, however, that imitation should not be seen as opposite to customisation. The idea of playing famous songs with an instrument created by the users themselves has been seen as a highly attractive possibility by several pupils. For example, one of them tried to play the soundtrack of the Titanic movie using two different sticks and a coin. Although the task was technically impossible to achieve, as the musical notes available were not the same as the ones of the song, the pupil expressed a sort of pride in showing his results to the classmate and wanted to teach him to do the same. Clearly, what the kid was proud of and wanted to show was not just the music he generated, but rather the particular way he modified the instrument to do so.

By the way, playing a large sequence of musical notes is a difficult task in Mogees and it is not what the instrument has been designed for. This aspect brought to the development of a simplified modality of playing called the 'song mode'. In this modality it is possible to preload a series of note in the form of MIDI file. The instrument simply steps through the different notes every time the player hits the surface (i.e., anytime an onset is detected in the input signal). It would be interesting in the future to evaluate such modality through another user study at primary schools.

5.3.4 Developing generic music skills

From an educational point of view, the feature of constraining the sounds produced to a given musical scale, or even to a given series of notes, opens up to the possibility to focus only on specific aspects of the practice, such as rhythm, while the sound produced being generically pleasant. This can allow pupils, and music students in general, to focus on other elements such as timbre, tempo and dynamics without feeling frustrated in making mistakes by playing wrong notes. Although skills such as rhythm, dynamic and posture of the gesture can be trained using a vast range of instrument (or we could say no instruments at all, just by tapping on a table with bare hands), Mogees can be seen as an interesting way to make the training more pleasant and satisfactory from the very beginning, as the experience in the school hinted. Moreover, the teacher noticed that, because the sound synthesis technique of the system has the characteristic of being strongly representative of both the dynamic, the tempo and the timbre of the gesture performed, the system offers a clear auditory feedback to the student about the qualities of his practice.

It would be interesting, in the future, to extend the software application in order to add more guidance in the learning process of specific music skills. For example, regarding the case of rhythm, it could be possible to evaluate the precision of tapping at a given tempo, providing scores and objectives, somehow similarly to the popular game *Guitar Hero*.

5.3.5 Instrument-specific skills

Some of the skills required to play Mogees, such as rhythm and coordination, are common in many musical instruments. Thanks to this, the system can therefore be seen as a tool to teach generic music skills. However, the Mogees practice also involves skills that are more specific to the instrument itself and can therefore be studied for such a purpose. The study with the children highlighted at least three skills that ca be seen as specific for Mogees.

The first skill consists in the ability to select the set of exciters to use. Such a selection defines the basics of the interaction with the instrument, the range of notes that can be triggered and the sound timbres. The second skill is the one of using such exciters to perform the same gestures reliably. Like in any instrument, being able to perform coherently the same gestures is vital to a performance. In Mogees such gestures can be quite different from the ones of any other instrument. For example learn to rub a coin on a table was a gestures that several children attempted to perform during the study and such gesture cannot be *recycled* from the training of any other instrument. A third skill that has been highlighted in the study was the ability to use the *capture* function. Such function allows to easily define the resonance properties of the instrument and is obviously very specific of the design of the instrument.

The study highlighted that such skills are not trivial and they all require time to be acquired. As mentioned previously, Mogees does not impose strict rules about how to study such skills and, on the contrary, aims to give a musical response to every possible action. It is therefore a responsibility to the player to explore different exciters and learn how to use them to trigger notes and capture new sounds. Interestingly, one pair of students, when asked with who they would like to play Mogees, answered '*with a proper musician, someone who has practiced*', highlighting the desire for a more clear guidance.

Whereas such a freedom builds the basics of the instrument and the goals of this research, encouraging users' customisation, and therefore it should not be lost, it would be interesting in the future to explore ways to guide the users to track their abilities against such skills. For example, it could be possible to add a *training* modality in the app. Such modalities could ask to the users to perform specific tasks such as triggering specific notes, creating repetitions of similar sounds over time and create timbres similar to given templates.

5.4 Contributions

This short, qualitative study hinted that Mogees is a simple and engaging tool to produce immediate satisfaction to the user whereas giving room for training and improvement. All the children were enthusiastic about using it for the whole lesson and some of them even asked for how they could download the app straight away. The study suggests that Mogees has the potential to support children' curiosity and natural interest towards sound and music improvisation even at a beginner stage. The instrument enables children to generate a wide range of timbres, as well as other musical elements such as dynamics and tempo. This physical interaction is also coped with the possibility of using headphones for individual, private creative work, which the teacher noticed to be useful in a classroom context.

This section describes the lessons learned and contributions for broader research scenarios.

Reality-driven interaction enhances customisation and creativity The tangible design of the system has been embraced by the children, who focused almost entirely on the physical interactions with the objects rather than on the iPod. The study showed how children were able to customise Mogees and make it their own, by creating their own individual combinations of exciters and gestures as well as by using the capture function to create new resonance models. Harnessing and sonifying the wide range of expressive movements that children would already know and makes interacting with objects they are already familiar with, the system allows for sound improvisations from the very first lesson, engaging them and easing the learning process introducing them gently with the use of technology for creative expression.

These results proves that tangible interaction offers great learning potential in the music education domain as well. The proposed *reality-driven* interaction extends the concept of tangible by offering to the users the possibility of free exploration with physical objects and adapting its output accordingly, motivating them to engage in an active experimentation with the technology and through their existing knowledge about the real physical world.

If the system supports visibility, imitation is propaedeutic to customisation Although the system proved to give an immediate sense of reward to the pupils, this study confirms the results presented in chapter 4 and supported by Carroll ([Carroll, 2004]) regarding the needs of a clear guidance. Pupils creativity and imagination needed to be triggered by an initial demonstration of the range of the possibility offered by the system, similarly to the users of the study presented in 4.6.

Using Alan Dix's terminology, if the flexibility of the technology (*allow different interpretations*) is coped with immediate and predictable outputs (*visibility* and *expose of intentions*), then imitation results to be not opposite but propaedeutic to customisation, giving an initial hint for a wide range of ideas of potential personal uses.

Customisation evolves through the path for virtuosity The Mogees system is designed to deliver an immediate reward and customisation of the interface of the instrument and its mode of interaction. However, there are challenging tasks that require specific training, such as the ability to trigger different notes reliably. Coupled with *visibility* and *imitation*, however, this initial difficulty offers the potential for improvement and for virtuosity, encouraging a longer term usage of the instrument. Such a task, however, is not only dependent by users' perceptual-motor skills (i.e. performing gestures reliably), but it also depends on the particular choices made during the customisation process (i.e. which exciters are used and how).

Wessel and Wright's *low entry fee with to ceiling on virtuosity* motto ([Wessel and Wright, 2002]) is supported by this design and applied to the topic of customisation. The property of enabling virtuosity through practice is here translated to the world of adaptable instruments, where the process of customisation, through being immediate, evolves and improve with experience over time.

Chapter 6

The technology in use: an 18-month collaboration with a music producer

Figure 6.1: Performance at Reggia di Venaria (Italy) for MTV Digital days, 12 September 2014. A custom-made structure have been created for the event, using 5 metal wires that covered a surface of 5x4m^2. Mogees has been attached to 5 metal wires, one chair and one window and performed by 5 dancers and by myself.

6.1 Introduction

The previous chapter described some important insights regarding the adaptable and reality-driven design of the Mogees system in the field of music education for young pupils. We move now to a radically different scenario, which resides on the opposite side of the spectrum for what concerns both musical and technical skills: a one-year collaboration with professional music composer and live performer Ed Handley from british experimental dance band Plaid [pla]. The technology is observed here in a more advanced stage of the customisation cycle, regularly employed in its practical real-world adaptation in the studio and in the stage.

Background Plaid is a duo composed by Ed Handley and Andy Turner, who published music under the Warp records label since 1991. They are considered amongst the founders of a music genre denominated by the media as Intelligent Dance Music (IDM), alongside their label mates Aphex Twin and Autechre. During their career, they also collaborated with icelandic artist Björk for the composition of the song 'Lilith' in 1997 [Not] and the consequent world tour. Their studio compositions and live sets are mainly based on laptop, MIDI controllers and iPads. When I first met the artist, the band was about to start writing a new music album; it was therefore the right timing for him to explore new music technologies for both their live performances and to use in studio production.

Chapter outline The research reported in this chapter adopts a *practice-based* approach: the work described here is practical and focuses on creating real, concrete artistic outcomes such as live shows, a music video and a studio album. Section 6.2 provides a qualitative description of the structure of this collaboration, dividing the workflow in different phases that happened iteratively. Section 6.3 integrates this description by illustrating the practical point of view of the first user of the system, myself. The personal and subjective strategies adopted during my practice for the live shows are described and motivated both from technical and artistical perspectives. Section 6.4 describes the artistic works that resulted from this first 18 months of the collaboration: a series of live performances and a musical video clip. Finally, section 6.5 discusses and generalises the work reported in this chapter and section 6.6 draws its conclusions in connection with the goals of this thesis.

6.2 The working process

These first 18 months of collaboration happened through various different phases that have seen the Mogees system being iteratively modified, applied and discussed. This section analyses this workflow by dividing it in different phases, described qualitatively through an interview which is fully reported in annex C.

6.2.1 Exploration

From the interview with the artist, it is very clear that at the beginning the main reason for him to start this collaboration was his interest for sound physical modelling: *'I've been interested in physical modelling since I first read about it (...) So initially it was the attraction to physical modelling, more of a sound generation thing than necessarily as a performance instrument.'* His familiarity and interest for this sound synthesis technique helped to understand the potentialities of the system very quickly. However, the artist immediately recognised the *exploration* of physical objects as an innate action in the usage of the system and he extensively used it as a tool to design new sounds, confirming the results reported in chapter 4. During the first weeks of the collaboration, we collected a series of 'found' physical objects ranging from small rocks and stones, to wooden and metal artefacts. Interestingly, rather than just collecting the objects, the artist also recorded a range of audio samples coming from different forms of interaction with the various objects. This library of sounds grew up during the whole length of the collaboration and has been used extensively in different phases of the work.

6.2.2 Software development

During the first weeks of work, the software has been modified several times to correct bugs and add new features, and every session with the artist has been seen as a deadline for the software development. This process, however, should not, in my opinion, be seen as a user-centred design approach. It was rather the completion of the software and its natural migration from a research prototype to a more professional and complete form that could be used on stage and in studio.

In music technology research, there is often a big discrepancy between designing a software prototype to be used for controlled experiments and implementing it so as it can be used in real live and studio productions. Furthermore, the system will need to be embedded in the existing working flow of the artist so as to fit a very specific role. In these practical cases, any bug or malfunction can stop the whole production and, therefore, extra testing is required and every meeting with the artist becomes a deadline for implementation. Also, there is a series of standard features that are required in audio production software which are rarely implemented in research prototypes. These include presets saving, low (or at least constant) CPU usage, compatibility with most used software (as streaming of OpenSoundControl messages to *MaxForLive* in our case) and remote control. Although the implementation of these features is time-consuming, the realisation of a professional quality system contributes to research by providing what was needed for the goals of this work: observing and evaluating the system being used in a real-world context.

6.2.3 Fitting the system to the existing working environment

As the two phases previously described were progressing, the artist started to understand the potentialities of the system for live performances and to embed it in his own working environment: *'I think the instrument part of it became exciting once I'd seen it in action, (...) when I had the Mogees in front of me and I could actually have a play with it. I think that's when I could see the potential of it as a musical instrument'.* For the whole length of the collaboration, he used two versions of the Mogees system: the standard one for iOS and another one for the *MaxForLive* environment. The architecture of the two platforms are very similar, with the main difference that the second one uses the audio input of the sound card as input device, usually achieving a better sound quality due to analog pre-amplification.

The main difference of using Mogees as MaxForLive plugin compared to the iOS app, however, is the possibility to apply a chain of digital audio effects and control them separately. Therefore, the artist could embed the software in the environment he was familiar with, Ableton Live, and apply sound design strategies that made the musical output of the system to be very inline with the sonorities he is used to work with. All the parameters of both Mogees and the audio effects are made controllable remotely using an iPad application called Lemur, which is regularly used by the artist in his live shows.

6.2.4 The ‘sequencing’ mode

After 8 weeks of collaboration, the artist asked me to program a new feature of Mogees that we called the *sequencing* mode. This mode has been used primarily for the production of a musical video, as illustrated in chapter 6.4.

The *sequencing* mode consists in bypassing the analysis technique implemented in the system for triggering the musical notes explained in 4.2.4. Instead, the pitch of the synthesiser is controlled remotely by sending OSC or MIDI messages. In this way, it is possible to synchronise the pitch of Mogees instruments with the tempo of a song triggering remote messages from the sequencer so as it always corresponds to a pre-composed score.

Although the movements of the performers still need to be on time, this mode offers the advantage of letting them to focus on the time and the aesthetic of the movements without caring about triggering the right note at the right time. The impact of this choice in our general considerations about the system will be discussed later in the chapter.

6.2.5 Studio production

As illustrated in section 6.4, the Mogees system has been used both for live performances and for the production of the video. However, there has been also a third outcome of this collaboration: the employment of the system for the realisation of the new Plaid studio album, *Reachy Prints*. During this process, the artist decided to modify the system even more. Instead of using the system in real time as it is supposed to be used, he fed the audio plugin version of the system using pre-recorded audio material. Such material came both from the recordings of the physical objects he collected during the whole year of collaboration and using samples coming from other commercial libraries.

The advantage of this approach is very clear: by using the system offline, he could control the system much more accurately. Interesting enough, however, for him this step was not seen as an *hacking* of the system but its natural usage in the context of a studio production: his main motivation about using the system was no longer its visual and performative elements. The attention rather moved to the sounds produced, which he found to be peculiar compared to the ones produced by other physical models, as the sound engine was *excited* continuously over time through his samples rather than using discrete MIDI notes as it happens with other synthesisers. A discussion about the impact of this process in the broader usage of the system will be presented in chapter 6.5.

6.3 Strategies of play: a subjective perspective

During the collaboration with Plaid, I personally engaged as Mogees performer, stepping into the interaction with the instrument in first person. I summarise here a series of subjective strategies I adopted, which constitutes a brief and highly subjective collections of hints about using the system for live performances.

6.3.1 Space is not enough

One of the first challenges that I experienced performing with Mogees is to resist to the temptation to consider the system as a *tabla-like* instrument that can be controlled purely changing the spatial position of the touch. The system analyses the *sound* generated by touching the object. And this is, in turn, the consequence of several factors: the spacial position of the touch, the exciter being used and the type of gesture performed. Therefore, it does not have to surprise if changing only the first of these parameters won't have a too strong effect on the timbre of the sound generated.

6.3.2 Selecting the exciters

The selection of the exciters is a key factor as it has a direct impact not only in the visual aesthetic of the performance, but also in the sound timbre and in the range of notes that can be triggered. Using exciters with different resonance properties helps reaching a wide range of possibilities. For example, as a metal coin or key is likely to generate a high pitch timbre whilst a rubber generates a low pitch, it is probably a good idea to have both available at the same time. Also, as analysed in chapter 5, exciters have their own physical affordances. It is therefore important to consider that selecting a set of exciters with similar affordances (for example a set with all percussive tools) will drive to a more homogeneous interaction.

6.3.3 Objects complexity and illusion

The more the physical object is complex, the higher is the range of possibilities for interaction. Plugging the system to a large surface, like a table or a floor, offers the advantage of placing a high number of exciters on it and have them easily accessible in every moment. However, its flat and smooth shape will constraint the types of gestures that can be performed on it to a two-dimensional domain. On the contrary, an object with a more *tri-dimensional* interactive area affords tri-dimentional gestures to be performed. For example, in one of my performances I used a metal ladder. This choice spontaneously led to play its different steps adding a third dimension to the interaction. This also naturally offers to the audience a greater *visibility* of the movements of the performer.

Furthermore, associating the physical objects present on stage always with the same sound preset will increase the *augmented object illusion*, creating a stronger synergy between the object and the sound it creates. Conversely, if the same object is combined with different sound presets during the same performance, this will allow for a greater variations of sounds but the artificiality of the technology will be unveiled and the illusion will vanish.

6.3.4 A trial-and-error approach

When exploring a new physical object to play with Mogees, I adopt what could be called a trial-and-error approach. I immediately try different combinations of gestures, positions and exciters aiming to find at least 3 or 4 clearly distinguishable timbres that fit nicely one with the other and are reliable. If this is the case, I then write a quick note on a paper to remember the setup, before starting to exploring a different setup employing different exciters and gestures. 3 or 4 timbres are usually a good number as they can be easily remembered. Variety of timbres can then be extended simply by having more Mogees available at the same time.

This approach is usually fast and rewarding. Frustration usually comes with uncertainty : if the physical conditions are not stable, the same action can bring to different sound outputs. For example, an issue I experienced sometime is to prepare a live setup in the backstage of a concert and then having to move the objects just before the performance. Or, likewise, preparing the setup under different acoustic conditions that the one of the real performance (i.e. using headphones or having different volume etc). Although we could argue this is a common issue of many other musical instruments, the fact that the physical and acoustic context *directly* affect the performance of the system enhance this problematic considerably.

6.4 Produced works

This first 18 months of collaboration resulted in a series of live performances and in a musical video, as illustrated in this section.

6.4.1 Live performances

A series of seven live performances took place regularly in several cities across Europe for different festival and events. The concerts showcased very heterogeneous setups, where different Mogees units have been connected to all sort of physical objects, ranging from natural objects such as stones and wooden boards (figure 6.2) to bike wheels (fig. 6.3) to bespoke sets built on purposes for the shows, such as a big 4x5 meters metal wires structure (fig. 6.1) and a metal scaffold designed to spell the work 'Play' (fig. 6.4).

Mogees constituted the foundation of the concerts from both a performative and acoustic point of view. Several Mogees units, which include both the hardware sensor and a mobile device (Apple iPod Touch version 5), were applied to a series of objects, which varied from concert to concert. Standard contact microphones connected to the laptop sound card have also be employed.

The shows alternated mostly two main situations: improvisation and structured performance. The improvised parts were very much based on performing different physical objects using different *exciters*. These sections alternated short solo performances of myself and duo improvisations. The structured parts, on the other side, were based on precomposed audio and midi clips in Ableton Live and were controlled live by Plaid, who also focused on live electronics processing of the sound material I was generating using Mogees. These sections have seen Mogees mostly used through the *sequencing mode* paradigm described in 6.2.4. The live electronic consisted in applying real-time audio effects to the tracks generated by the various Mogees systems, which where routed to a main audio card, as well as generating and controlling the playback of rhythmic tracks and handling the overall audio mix, and was controlled using various MIDI interfaces and iPads running the Lemur application.

In two occasions the concerts included performers other then myself and Plaid, the first one seeing the presence of an electric guitar player and the second one with five contemporary dancers (fig. 6.1).

All concerts also included video projections and lights. These were different for every performance, varying from the projection to a screen of the live performance using video cameras (so as to make it more visible to the audience) and applying realtime video effects to pure generative video material controlled by the outcoming sound. Lights have been controlled by generating DMX signals in realtime based on the amplitudes and frequencies of the various Mogees.

Figure 6.2: Brunel Electronic and Analogue Music (BEAM) festival, 22 June 2012, Brunel university (London). Mogees has been attached to various materials such as wood, metal and stones.

Figure 6.3: IRCAM Live, 28 November 2012, Gayté Lyrique (Paris). Performing with a bike wheel.

6.4.2 The 'Elementary Excitations' video clip

After 4 months of work together, we decided to produce what resulted to be a 3 minutes videoclip entitled *Elementary Excitations*, successively published by Warp records. The video can be found at this link: `http://bit.ly/1u1XDOP`.

We worked with a total of seven performers and the concept of the video has been highly influenced by the ideas of customisation and *found objects*. Before starting to film the video, we had one preparatory session where we showed the Mogees system to the performers and ask to each one of them to improvise with the system individually. Then, on the day we shot the video, we asked them to select the sounds they wanted to control, both amongst the ones we prepared before hand and by looking around them, to find a physical object they would like to perform with and attempt to elaborate an original way to interact with it.

Figure 6.5 illustrated four different moments of the production. The first scene sees myself performing Mogees with a selection of everyday objects arranged on

Figure 6.4: Performance at the RoundHouse (London) for the Reverb festival, 21 August 2014. Mogees has been applied to a various objects including a metal scaffold, built on purpose to spell the word 'Play'.

a table, from kitchen silverware to coins, keys and stones. The second scene involves a dancer from the London Trinity Laban Conservatoire of Music and Dance performing a track of the piece using a metal staircase. An electronic music performer and another dancer decided to perform attaching the Mogees in a metal staircase, whose shape and structure reminded them of an harp. Finally, a team of break-dancers selected a wooden board that they found in a building site nearby, in order to dance over it and sonifying the sound of the steps. Although visually very interesting, this last scene has been removed from the final video and constitutes a separate videoclip for technical reasons.

The compositional process The compositional process of this piece is unusual and is worth to be briefly discussed. The aim of the video was to show a real interaction between the performers and the system, and therefore we wanted to avoid the standard approach to music videos based on performing *in falsetto* (i.e. mimicking the performance of a sound which is actually pre-recorded). In-

Figure 6.5: Different performers have been asked to select an object to play with and invent their own way to perform. From top-left clockwise, Mogees has been applied to a table, a radiator, a staircase and a wooden board.

stead, we let the performers to improvise and we successively edited the video material so as to build the song *from* it.

Before doing the video recording, we prepared a series of audio track and a rhythmic audio loop. Each audio track was represented by a Mogees sound preset and a score of notes distributed on the time line of Ableton Live. We let the performers selecting one of the tracks. We then diffused the playback of the audio loop and ask them to perform improvising accordingly. The sensors they used were connected to the Mogees preset associated with the track they selected, and the pitch changed accordingly with the score. We recorded an average of 15 minutes of free improvisation for every performer, which served as material for the composition of the song. However, instead of recording the audio output of the Mogees, we recorded the unprocessed audio signal of the sensor. This left us with the possibility of making changes to the Mogees preset and notes afterwards.

6.5 Discussion

This collaboration allowed to closely observe the strategies adopted by a real-world artists while using the Mogees system in order to accomplish his artistic tasks. This section provides a discussion about what has been observed qualitatively during this period of work, connecting the observations with the results presented in the previous chapters of this thesis.

6.5.1 Exploration is the point of departure

Chapters 4 and 5 showed how Mogees has generally been seen as a sonic exploratory tool for radically different purposes. The experience reported in this chapter firmly confirms this vision. Describing his way of using the system, Ed Handley refers to exploration as the spontaneous action someone would do when approaching the system: *‘ You want to see how [a physical object] propagates sound. How it vibrates. You just want to see how it sounding through Mogees. (...) It encourages you to look at the objects around you and think about them, or discover them in an acoustic sense ’*). In the interview, he stresses on the timbre of the sounds and the way they are designed through a physical process: *‘the basic attraction of physical models obviously is how you excite them, so you've got this two sides to it; you've got the sound that's actually generating, but how you stimulate the model is just as important, and I think for me that's why it was attractive as a type of synthesis (...) With Mogees, the modulation is physical. The modulation is something very distinct, in that it's the exciter, it is what you chuck into the sound.’*, stating that the sound design starts in Mogees from this physical exploration phase, by selecting resonators and exciters and associating them with different physical models.

However, the way this phase fits in the composer's working process is radically different from, for example, the one of the music educationalist. While in music pedagogy the action of exploring physical objects and understanding their properties is the final goal of the education process, for the composer this phase served as an environment for sketching ideas. And the result of this process has been the development of a series of physical objects and sounds to be used in the production, rather than the experience itself. The exploration phase has not been used directly in the final outcome of the work, but as a preparatory process for then building a more controlled environment that could reflect compositional rules and specific music decisions.

6.5.2 Master what you need, forget the rest

The request to implement the *sequencing* mode clearly showed how the attention of the composer was not in the possibility of improvising with the system and creating melodies. He explicitly asked to modify the system so as to bypass the gesture analysis and control the system with the tool he already mastered: the sequencer.

At first instance, this request corresponded to me to a migration to a new different system, as what I, as *designer*, judged to be the main focus of the system was simply removed. However, from the point of view of the composer this modification has not even been seen as an *hacking* of the system, but rather just as a secondary *mode* of using it. He still claimed that the system was letting him doing what he wanted and then it was just easier to work in this way.

This shows the reluctancy in changing his existing and well established modus operandi. Rather, this modification allowed him to entirely focus his attention on the features of the system he was interested on and would not be able to achieve otherwise: creating novel sound timbres and adding performative elements to the music.

By exciting the audio physical model using audio samples, in fact, he was still able to obtain sounds that were not achievable by using the other approaches he adopted previously: *'The modulation is something very distinct, in that it's the exciter, it is what you chuck into the sound audience often likes to see something happening, and they like to see something physical happening.'* Likewise, he felt that in the video production the relationship between the movements of the performers and the sound created was still very strong, even if the 'note triggering' feature was missing.

This approach is somehow different from the one adopted by the music educationalist, for who this feature constitutes an important challenge of the system in the long term. As discussed in chapter 6.3, the view of this producer is also very different from the one I had in mind when I designed the system itself.

6.5.3 Delegated performance

'The creator of a sound object, instrument or performance has an informed position. Through delegating the performance to a non- expert, cliches can be avoided and the idea of naive and authentic performance can be taken further' [Richards, 2013].

This idea of *delegated performance* defined by John Richard offers interesting cues about the consequences of customisation in performative art. During the realisation of the Elementary Excitations videoclip, the performers were forced to select their own interactive system and design their own strategies to play. This adaptation naturally encouraged the performers to adopt an aesthetic more directly influenced by their cultural and technical background. This effect will be limited in a context of free improvisation with a more strict musical instrument, as the interaction would have been constrained by the rules of the instrument itself. The work *Kontakthof* by Pina Bausch constitutes a notable example of this approach, where only amateur performers over the age of sixty-five have been selected for their inexperience in performing.

6.5.4 DIY virtuosism

This work offers interesting analogies with the world of Do-It-Yourself (DIY) musical instruments, both at the hardware side such as with instruments built for example using Arduino, and at the software side with instruments built using high-level programming languages such as MaxMSP or PureData. It has been observed how the design and the development of DIY musical instruments is an inner part of the creative process[Richards, 2013]. Therefore, the affordances of the instrument and the way it can be played are constantly evolving and the idea of practice to master the instrument so as to become a *virtuoso* player loses its focus. This idea of evolving instruments has been broadly discussed in the NIME community, proposing standard techniques for the evaluation of new digital music instruments [Orio et al., 2001] and claiming the need for standardisation as a requirement for the development of novel form of performance practice.

With a system like Mogees we can observe a similar phenomenon, although the design process is much more immediate. Everytime the system is applied to a different object, the rules of the game change and performers need to explore and rethink their strategies again. Moreover, because these rules can be reinvented, every attempt to virtuosism in the performance is likely to be affected by the artists' own styles rather than be peculiar to the instrument itself, as we witnessed through the work with the dancers (6.4.2).

We can observe however that another form of *virtuosity* is allowed in the exploration phase. It is in this phase that, like in DIY instruments, creativity and originality can be expressed through design, rather than through performance: the conception and realisation of the instrument and the interaction themselves is already part of the artistic process.

6.6 Conclusions

This chapter detailed the different approaches that seen Mogees being used during practical real-world artistic productions. By observing the technical workflow phases and the artistic achievements that have been reached during this long period, it discussed the very personal strategies adopted by the different persons that have been involved in the process: the music producer, the performers of the videoclip and myself, the design and first user of the system.

The flexibility of the proposed design has been reflected in its different usages, where the producer ended up controlling the system in different ways for the live shows and the studio productions, and the dancers expressed their ability in reusing their pre-existing performative skills to employ the system in different ways: breakdancing on the street, playing a staircase miming the movements of an harp and dancing next to a radiator. We witnessed a *reluctance* by the producer (and confirmed by the performers) in changing his own artistic modus operandi, and rather a natural attempt to adapt the system itself to fit their work space. He changed gradually the system bypassing the features he was not interested in and enhancing the ones he was interested into and that could deliver to him something that could not be achieved in his existing working environment.

These approaches radically differ from the ones of its designer, which happens to be the first performer of the system. The *imitation* process, i.e. myself demonstrating a specific using of the technology (tapping on a table as if it was a percussion) guided the users but didn't stop them taking advantages of the flexibility of the proposed technology to adapt it to their skills and workflows. Such a result confirms what has been found in the studies described in chapter 4 and 5 regarding the reality-driven interaction to be a compelling design to boost adaptation and creativity, as well as imitation to be propaedeutic to this goals.

Chapter 7

Conclusions

7.1 Developmental narrative

I want to move back for a moment to four years before the moment I'm writing these conclusions, when I decided to write a PhD in music technology. The motivation that brought me to undertake this journey is probably the same as many of my colleagues in the community of New Interfaces for Musical Expression (NIME). We simply love the idea of waking up in the morning and spending the whole day in designing and experimenting new musical instruments. I experienced a dawning realisation that within this community we tend to identify ourselves not just with the sounds we produce but with the instruments we create to that end. There is a genuine feeling of excitement every time we think about applying existing techniques together in unique ways so as to build a new interactive music technology. As the time passed, I realised that what I wanted to do was to share this feeling with people outside our relatively small community, to empower them with a sense of ownership towards the musical instruments.

Three years later I travelled to New Delhi for a speech and I had the occasion to show the Mogees system to an Indian Kathak dancer. When I watched her placing the unit onto a table, jumping onto it and starting to dance, triggering different notes with her feet in ways I could never have been capable of, I saw the cultural reach of the instrument had taken on a life on its own.

Reflecting on the journey between my impetus and watching the Kathak dancer, I can recount many challenges, opportunities, decisions and discoveries that I would like to share in these conclusions. At the beginning and later during the development of the project, I had serious doubts about whether it is possible to create a shared universal experience of adapting new musical instruments. The more parameters we expose in an interface, the more the technically expert users can tune the instrument as they want but the less the system will be appealing to general adopters.

We can observe that very often in NIME performances, a big chunk of the audience enjoying a show belongs to the very same community and share the fascination towards the instrument as much or even more than they do with the music that is produced, considering the instrument to be a real artistic outcome. I was curious whether or not this could have a wider and popular appeal.

Before undertaking this PhD journey, I worked for three years on the Gesture Follower technology under the supervision of Frédéric Bevilacqua at IRCAM. I soon realised that during those years I acquired a good deal of expert knowledge on using it and spent a lot of time making this explicit to other composers and performers. I wanted to provide enabling conditions for them, thus removing myself as a necessary step in their workflow. The next step has been to realise how diverse the users base of the Gesture Follower was, and how their requirements were different and specific to the various individuals and application scenarios. The potential of this technology was not purely in the precision of the algorithm when employed correctly, but in its adaptability to different contexts and applications.

After this experience, I wanted to preserve the essential qualities about adaptability and realtime continuous feedback offered by the Gide system, which are achieved thanks to the features provided by the Gesture Follower, but moving toward a more accessible and portable instrument. I was fascinated by the possibilities of embodiment and cohesion between the interface and the sound generator without mediation through a computer and thus decided to embrace the field of Tangible User Interaction. From within this field, I leveraged the existing body of work regarding embodiment, customisation and appropriation of technologies, thus leading to the trajectory towards what ultimately became Mogees.

7.2 Potential academic impact

A commonly asked question within the NIME community is how to step beyond an academic prototype to an instrument with popular appeal [Newton and Marshall, 2011]. As Jordà perfectly summarised in one sentence, *'Many new instruments are being invented. Too little striking music is being made with them'* [Jordà, 2004a].

Although popularity might not necessarily be a good indicator of the importance of any research undertaking (and can depend on stylistic choices, marketing strategies and pure chance), popularity allows us to observe the system being used in practical real-world situations.

Tahiroglu and Farnell [Farnell and Tahiroglu, 2014] compared NIME to a playground of opportunity, which is the charm of the field, but it is a restive one (in the sense of De Tocqueville) which sometimes lacks a well defined *telos* and has an ever changing set of evaluation criteria. As one continuously *chases after The New*, much potential depth of investigation is left behind. I believe appropriation is a fundamental feature in a musical instrument, and in order to be observed and measured the instrument needs to be watched while in use.

An extensive minutiae of activities such as graphic design, tutorials and instruction manuals are essential to introduce an instrument to a wider audience. These are too often forgotten because they do not seem directly relevant to research goals. If a new system is proposed every year, these series of small tasks are harder to achieve, forcing instruments to stay at the prototype level and real world usage is not explored. Users, and not designers, manage the coupling with a technology and a design process can never be completed without it.

In some senses, Gide and Mogees give us insights about the potential of customisation as an inbuilt feature of an instrument that sharpens the focus on its uptake by a broader and diverse range of audiences. Hopefully, my intended contribution is a widening of the interest beyond the small community of NIME in the design of unique instruments as an art form and I believe the way to achieve that is to create the conditions for our prototypes to leave our research studios and see the light of other performers stage.

7.3 Potential industrial impact

As evidenced by the initial uptake of 1600 customers in a crowd-funding campaign, I believe the technology behind the Mogees system described in this thesis demonstrated real potential for mass market. The low cost of hardware manufacture makes this technology appealing and affordable to a broad range of people, from traditional percussionists to electronic music producers, gizmo lovers and music teachers.

Specifically in music education, the relevance of this type of system is supported by the contemporary English National Curriculum for 7 to 11 year old pupils excerpted below from [ukn, 2013], which drove me to the experiments presented in chapter 5:

> *Pupils should be taught to:*
> - *identify how sounds are made, associating some of them with something vibrating;*
> - *find patterns between the pitch of a sound and features of the object that produced it;*
> - *find patterns between the volume of a sound and the strength of the vibrations that produced it.*

7.4 Potential artistic impact

This thesis seems clearly situated in a long line of the reality-based art movement. Exponents of this include Marcel Duchamp, who at one extreme simply appropriates existing objects wholesale, to a visual artist like Max Ernst, who incorporates the notion of *frottage* as technique where reality strongly inspires and is present within the artistic outcome. My work continues in this tradition of reality-based art adding a significant contribution to it in its musical equivalent.

Formalising the concept of found objects as musical instruments raises the value of those unique objects as art pieces in their own right. I hope that this work can provide insights to legitimise the value of unique instruments as artistic outcomes beyond the NIME community towards a broader domain.

7.5 Responses to some critical objections

During the last few years I have been showing my research on many occasions and had the opportunity to engage in constructive debates that often aired similar topics. It is therefore useful to summarise in the conclusions of this thesis these objections and my responses to them.

Musical instruments should not be immediate, otherwise they prevent deeper forms of interaction. I believe the premise of this argument is faulty. Studying techniques to ease the early usage of a system does not preclude depth of interaction. Rewards can be provided in an initial phase while still ensuring room for further and deeper rewards arising from more persistent exploration.

Are you implying that rigid interaction paradigms are inferior to adaptable instruments? I believe this is a false dichotomy. Whist they are different interaction paradigms, each simply enables different appropriation cycles. A rigid interaction paradigm such as that of a piano offers different challenges to appropriation than in an instrument that is designed to encourage it. Both offer their own kind of value to exploring performance and this phenomenon simply happens during two different phases, the former after several years of practice and the latter from the very first day.

The range of timbres produced by the Mogees instrument is very limited. From a certain perspective this is indeed true since all the sounds are strongly imbued with the characteristics of the resonator based synthesis. However, one must not confuse the spectral range of timbres with the nuanced diversity of those timbres. For example, one can consider the range of timbres produced by a violin as extremely narrow by the same philosophy. Perhaps with usage and familiarity, somebody who raises this objection would discover the potential for the subtlety in the Mogees timbres.

The algorithms employed for the implementation of the Mogees system are not novel. Not forgetting that the very essence of NIME research lies within applied science, we can observe that any algorithm can be a combination of smaller ones. Thus, the uniqueness of the Mogees system resides in the combination of other familiar design patterns to achieve novel research goals. I believe that the novelty value resides in the overall system and in the combination of its parts.

7.6 Implications for future work

Although the current implementation of the two systems proposed is satisfactory and serves the purposes of this research, there is room for improvement in several technical areas. For example, for the Mogees system these include:

- Extending the interaction paradigm proposed to supervised machine learning approaches in order to increase the number of gestures that can be recognised.
- Using the same data to drive more complex synthesis systems for sound generation.
- Extending users' input to the use of multiple sensors.

Music technology is a very challenging research field, requiring continuous and realtime interaction and a high expressivity bandwidth and addressing a wide range of users with different needs. It is therefore a good testbed to develop wider application scenarios, capable of producing insights valuable to the original HCI domain.

It would be interesting to extend this research to the question of how to define a 'quantifiable' measure of customisation both as an intended feature of a system and as a practical phenomenon that occurs during the user experience. Furthermore, it would be worthwhile to study its *perceived* value through a series of comparative studies.

7.7 Our technology

The design cycle is never completed by a designer, it always involves the users. Users will always adapt and customise technologies to fit them into their working practice, sometimes even challenging the original intentions envisaged by its designers, and this study shows how this process can usefully be encouraged and magnified. The advantages of technologies that take this process into account are twofold. On one end, they let users focus on their objectives rather than on adapting themselves towards the new technology. On the other end, they provide vital information to the designers allowing them to better accommodate users' needs.

Hopefully this work highlights for the reader the importance of legitimising free forms of experimentation with technology. Technologies should inform us about the available possibilities they provide and drive us to adapt them to our own unique interpretations, without a fear of violating rules imposed by the designer and enabling the technology to be truly *ours*.

Appendix A

Appendix: HMM procedure

As described in [Rabiner, 1989a], the forward procedure can be used to estimate the probability distribution of a sequence of observation $O_1, O_2, ...O_t$. This requires the computation of the $\alpha_i(t)$ variable which corresponds to the probability distribution of the partial observation sequence until time t, and state i . It is computed inductively as follows:

Initialisation

$$\alpha_1(i) = \pi_i b_i(O_1) \qquad 1 \leq i \leq N \tag{A.1}$$

where π is the initial state distribution, and b is the observation probability distribution.

Induction

$$\alpha_{t+1}(i) = [\sum_{i=1}^{N} \alpha_t(i) a_{ij}] b_i(O_t) \qquad 1 \leq t \leq T-1, 1 \leq j \leq N \tag{A.2}$$

where a_{ij} is the state transition probability distribution.

From the $\alpha_i(t)$ variable we can compute two important quantities:

1. Time progression of the sequence, related to the recorded example

$$time\ progression\ index(t) = argmax[\alpha_t(i)] \tag{A.3}$$

This last value can be alternatively estimated by the mean (expected value) of the distribution $\alpha_i(t)$

2. Likelihood of the sequence.

$$likelihood(t) = \sum_{i=1}^{N} \alpha_t(i) \tag{A.4}$$

This quantity can been used directly as a *similarity measure* between the gesture being performed and the recorded reference. Other similarity measures could also be derived by combining the *likelihood* and the smoothness of the *time progression* index.

Appendix B

Appendix: Additional data about the Mogees evaluation at primary school

This appendix integrates the information about the user study described in chapter 5, including the documents provided directly from the music teacher Nancy Evans. Section B.1 presents the background questionnaire filled by the pupils before the beginning of the study in order to assess their technical and music preparation. Section B.2 provides a summary of the comments of the children observed during the study. These first two sections constitute an overview of the Mogees system from the pupils point of view.

Section B.3 then presents the report from the music teacher, developed after a series of discussions with myself and after the analysis of the video recordings, and offers the perspective of the teacher herself regarding the usage of the system in the context of her work.

B.1 Background Questionnaire

- Do you use computers/iphone/ipad/tablet etc at home or at school?
  - Yes (18)
  - No (0)
- How often do you use the computers/iphone etc/Wii/gameboy?
  - Every day (3)
  - Most days (3)
  - Sometimes (12)
  - Never (0)
- What do you do when you go on the computers/iphone etc?
  - Play games (18)
  - Watch DVDS (13)
  - Look up thing (14)
  - Wii (12)
- Have you ever made music using a computers/iphone etc?
  - Yes (10)
  - No (8)
- What programme(s) have you used?
  - Can't remember (2)
  - Karaoke prog (3)
  - Singing Mario (1)
  - Bruno Mars Lazy (1)
  - Mini piano (1)
  - Make Music 95 (1)
- Do you play a musical instrument?
  - Yes (18)
  - No (0)
- Have you ever created your own music before?
  - Yes (18)
  - No (0)

B.2 Summary of children's comments about Mogees

What does Mogees do? *The children described Mogees as being able to make songs, sounds and music. They said that they could make and create their own music. They described playing it by tapping, banging, scraping, thumping and hitting and that they could change sound or notes by using different objects ('stuff from round the house') or changing the sounds on the app. One child suggested that lighter objects made quieter sounds and big objects made big noises. The children offered different explanations as to how Mogees worked:*

- *'It takes sounds and makes an even better sound from it'*
- *'Makes music and sounds using the microphone.'*
- *'It's a radioactive speaker. It picks up waves - radio waves.'*
- *'It makes echoes when you bang on the table.'*
- *'Makes the sounds on the table go into the iPod and then the speaker.'*
- *'Takes the sounds of stuff.'*
- *'Picks up the sounds and goes into the iPod. Makes the iPod make a different sound.'*

One child described tapping the table to get sounds using different parts of the hand as being like using a drum in Pakistan. When asked whether he meant the tabla drum he agreed and mimed with his hands. Another child described it as 'making patterns with fingers.' Favourite exciters included: Pickle onion spoon (4), ping-pong balls (3), chopstick wooden knife, paper clip, black metal disc.

What do you like about Mogees? *The children were overwhelmingly positive about using Mogees to make music describing it as making nice songs, music and sounds. Positive features mentioned included being able to: make their own music; change the sounds using different objects; and, capture and listen to their own sound. They liked the simplicity 'all you need is a bit of cable and stuff to do it'. Quite a few children mentioned enjoying making loud sounds/noises:*

- *'It creates brilliant music that you want to play not music that you don't want to play.'*
- *'If you don't like other music [you can] make some music you do like.'*

One child suggested that you could use the music created to go with animations.

Was there anything you didn't like? *A few children mentioned that the sounds were too loud. One child mentioned that when they wanted it to stop it kept going and another said she didn't like that sometimes you could still hear your noise [noise of playing object/table] over the sound of Mogees. One child didn't like the 'strange' names given to the sounds.*

Was it easy to use? *The children said they found Mogees easy to use. Difficulties mentioned included:*

- *specific exciters/exciters not working ('scraping the metal thing it didn't work', 'the brush didn't make a sound on the table');*
- *not knowing what sound an object was going to make before using it ('every time you do something you don't quite know what sound it will make');*
- *trying to find the sound 'you were looking for'.*

All children said that they thought the difficulties would improve with practice and when 'they had tested them'. One child mentioned the practical problem of not being able to hear your bit when there were two people playing at the same time and with different objects. Again, the child agreed that this would improve with practice. One child mentioned having problems getting the contact mic to stick her chosen object.

Is there anything you would like it to do that it can't do at the moment or suggestions for improvement? *The children suggested being able to:*

- *Use earphones with it;*
- *Make cat or dog noises;*
- *Use their music as a ringtone;*
- *Make tunes, other music, proper music;*
- *Make more lighter sounds;*
- *Take sounds and make them into words;*
- *Use guitar or drum sounds;*
- *Tune it into a piano;*
- *Use high and low (pitch);*
- *Send their sounds to another iPhone which has the same app.*

What object would you like to try with Mogees, how would you play it, when and with whom? *Some wanted to put Mogees properties to a functional use, e.g. sticking it to the bedroom door so that they could hear someone entering or sticking it to the cat so that they could tell when it was hungry by the app picking up its grumbling stomach. Others thought about objects which themselves move or are moved by the wind: sticking to a car that goes over bumps; an electric toy that moves; a ball that could be kicked, thrown or bounced; or, a tree that blows in the breeze. Some thought of objects which has interesting sounds: a bottle with someone blowing over the top or a resonant biscuit tin. Three children suggested attaching it to something that already made music: an iPhone and the TV and wanted to capture songs they produced. Two children wanted to capture the sounds of a beating heart. One had the idea of attaching it to a gate and using their dogs paws as exciters. They wanted to play Mogees inside, outside, at school and at home. Mostly the children named their friends or member of their family as the people they would most like to play Mogees with. However one pair said they would like to play with a 'proper' musician - 'someone who has practiced'). 'They would know the sounds to make and you could add sounds'. This question was a little unclear as to whether it referred to an object to capture the sound of or an object to attach the mic to and play. The answers also reveal the misconception that it is the sound that is being captured rather than the vibrations of an object. Even when this was explained during the experiment I'm not sure it was understood.*

B.3 Report from the music teacher

B.3.1 An initial assessment of the strengths and weaknesses of using Mogees with young children

Strengths *Mogees is simple to use and produces immediate sonic satisfaction to the user. All the children were enthusiastic about using it and some wanted to know how they could download it immediately. Mogees has the potential to support for children's natural inclination towards multimodal as well as music play/improvisation. For non-instrumentalists there is a wide range and diversity of sounds open to them even at a beginner stage. Mogees allows for a range of timbres/sounds to be available to the children through one 'instrument' and therefore their improvisations explore the use timbre as well as other musical elements such as dynamics and tempo. The children can create more complex music with Mogees than they would be able to as beginner instrumentalists as it harnesses and sonifies the wide range of expressive movement gestures that they already make. The children are able to appropriate Mogees and make it their own by create their own individual instrument with a unique sonority(ies) through the capture function, giving them ownership over the sound as well as their improvisations. Mogees utilizes simple technology that many children already have access to. Mogees doesn't require special beaters as many percussion instruments do - in fact it thrives on using every day objects as exciters - objects found in any child's home. Traditional beaters can only do limited things - exciters, individually and collectively had more possibilities. Mogees software and hardware, unlike many music technologies, focuses the children's attention on the table/instrument and on music making rather than the iPod. The potential of using headphones for individual creative work in a classroom context is very useful. Mogees connects well with children's worlds that increasingly include the use technology for creative expression.*

Weaknesses *Children traditionally learn to play and improvise on instruments by being able to find sound images/ideas over and over again. This process is frequently aided by the architecture and structure of the musical instrument they are playing. This is more difficult with Mogees as it is not spatialised. The* *children will need time develop other, yet unknown, strategies for doing this. It is the ability to do this that allows them then to return to previous idea later in an improvisation or to transform/vary the original idea. This is an essential requirement for children creating improvisations and compositions with coherent and thought out structure. Further study as to what strategies children might learn/develop, or be taught, is needed. While Mogees is excellent for individual and pair work in the classroom with headphones, as with any classroom group music activity, there is the issue of sound pollution and the difficulty hearing individual or small groups musical ideas when multiple individuals or groups are working. Also if Mogees were to be used in a classroom context there would be issues with leads (microphone and speaker) and difficulties separating the speakers so as not to get feedback. It is difficult for beginners to achieve a full range of pitches. However is it likely that this would improve with practice and with better understanding of how Mogees works. Further study is needed to explore this. Something that also might be considered is, given that Mogees is such a radical new way of making music, how much should it try to replicate conventional instruments with its use of traditional tonal scales which reinforce western notions of what music is, rather than open up new sonic possibilities which are less pitch focused or use conventional tonalities.*

B.3.2 How did the children interact with Mogees and what children's Mogees music sounds like?

Jo Glover talking about children's early composing and improvising with instruments says:

> *'Any instrument suggests its own musical structures'. The musical patterns young children make with instruments often arise from a response to the visual structure and the action and sound patters which these suggest.'*

For example:

> *'Pitched instruments or keyboards with notes presented in order - low pitch to high pitch - offer a visual analogue of what is heard and this becomes very important for children in making the transition from random to structured melody making.'*

What is interesting about Mogees is that it does it not offer the children a visual analogy, as pitch is not spatialised. Instead it requires different way to approaching it as an instrument and learning how to play it. This lack of visual analogy or lack of architectural clues could present the children with the problem of re-finding ideas or guessing at what sounds particular gestures/actions will produce. As one the children said 'every time you do something you don't quite know what sound it will make'; and another expressed difficulty 'trying to find the sound you were looking for'. What we don't know at the moment is what strategies children will use, without recourse to visual clues, to learn to master the sonic possibilities of Mogees over time. Further study as how the children might do this and how adults might support them is needed. Those children whose comments these were all agreed that they got better at finding what they wanted with a little bit of practice.

Jo Glover also comments:

> *'At this stage, improvising and composing are wholly dependent on the skills of producing and sounds from the instrument. Trying out new techniques, exploring sound quality, listening to oneself play, repeating over and over, building up new musical ideas and revisiting music made before al require time alone.'*

This is a very useful reminder that Mogees like any conventional instrument, despite producing what can seem like immediate success/satisfying sounds, needs practice. Similarly, when observing young people using DubDubDub (a digital instrument which allows the user to remix the sonic content of the Internet, in real time) [Savage and Butcher, 2007], Savage and Butcher noted that learning to control the instrument could not be short-circuited. What Mogees does allow for, unlike traditional classroom percussion, is the opportunity to practice alone in a classroom environment using headphones, thus preventing one of the biggest problems of classroom music - noise pollution and the inability of the children to hear what they are individually doing. Despite the rectangular shape of the table, none of the children attempted to use the table as a piano or other keyboard. This suggests that the children were not superimposing onto the table the architecture of a known instrument but instead approaching it as a neutral space on which different kinds of actions/gestures could be played as suggested by the exciters themselves.

With Mogees set up as in this experiment - table plus various exciters - it could be said that it is the 'actions' available or affordances of the different exciters were what the children explored rather than the architecture of the table. What is important to look at is whether when using Mogees the children use the exciters in their conventional way/function e.g. knife cutting, ball dropping or transcend this to use them in order to find the best, most interesting sounds. Sometimes of course the traditional way of using and object/exciter/hands may produce the most successful sound too. In this study there were examples of children transcending (or repurposed with musical intention) the conventional use of the object/exciters to find 'better' sounds. Sometimes, ways of playing one exciter conventionally led to a second object/exciter being used in the same way but which was unconventional for that object/exciter.

One child perceptively commented that playing Mogees was like a 'drum in Pakistan' and mimed actions. He was referring to a table drum that is played with different parts of the hand producing different sounds - possibly one of the closest analogies to Mogees along with a prepared piano. This offers up the question as to whether Mogees is best used with familiar objects - which is very appealing because it is accessible without special kit or with specially designed exciters which are less likely suggest particular ways of playing and might be optimized to initiate the best range of sounds from the Mogees software.

B.3.3 Ways of playing

Tapping was the most frequent method of using the exciters. Some children only tapped but many children went much further. Both from watching the videos and analyzing the children's comments, the pickle stick came out on top as their favourite exciter. This was followed by other 'beater' like objects such as the paintbrush, chopstick and wooden knife. A preference for beaters over hands, common when using drums with children, was replicated in the study. However, after the demonstration, not only did the children use their hands more they also used more of the non-beater like object/exciters. The full range of techniques to use individual exciters included:

- *Pickle stick: tapping, twisting, scraping, rocking on edge of table.*
- *Paintbrush: tapping, brushing (mostly unsuccessful unless done on the mic),*

rocking on edge of table.

- *Coin: dropping, scraping, tapping.*
- *Wooden knife: tapping, cutting, scraping.*
- *Ping pong balls: rolling, natural bouncing, controlled bouncing by either rhythmically catching and dropping it or creating rhythmic bounces using on the flat of their hand, putting the ball inside a pot and spinning or shaking the pot on the table.*
- *Metal black disc: scraping, spinning, dropping, tapping.*
- *Paper clip: tapping.*
- *Green plastic leaf: scraping, flapping, rubbing and tapping.*
- *Grey plastic bowl: tapping both ways up, sliding, spinning, dropping a ball in.*
- *Hands: fingers tapping, knocking, flicking, scratching.*

Though traditional percussion instruments require and afford different actions in order to produce a range of sounds, Mogees offers a much bigger and more diverse set of actions/gestures within one instrument. Likewise, though you can use different kinds of beaters to play drums and glockenspiels etc. the possibilities of individual beaters is limited whereas using Mogees there is a wide range of affordances available from the beaters/exciters and ways of using them, collectively and individually as part of one instrument.

B.3.4 Children's Mogees music

The children's music involved:

- *Repeated taps with one exciter - sometimes moving onto a new exciter to do the same thing with the same hand, sometimes exchanging for a new exciter.*
- *Tapping ideas with two exciters, alternating to create a repeating pattern - sometimes this transferred to other different pairs of exciters or hands (with the hands this sometimes, but not always, meant doing something different things each hand e.g. a flick and a knock).*
- *Repeating rhythmic patterns where the two exciters are played together with the same idea.*
- *Patterns where the two exciters are being played at the same time but each*

hand is doing something different. E.g. tapping in one hand with a chopstick while dropping a coin with the other.

- *Patterns that are played once on one object then transferred from one exciter to another.*

Some children's play with Mogees could be categorised as 'testing' mostly without obvious musical intent. E.g the exciters were picked up one at a time and tapped or dropped with no other musical features. A few children's included variations in tempo and dynamics within their improvisations and used silence expressively. One child, probably the most musically experience of the group, said that when playing Mogees he had been recreating some music he already knew, one piece he couldn't remember the name of and the theme tune to The Titanic.

There was a wide spectrum of musical engagement and skill demonstrated by the children using Mogees in the study. From 1. being the highest level of skill and musicality to 6. being the lowest level of musicality and skill the following spectrum was observed within the study:

1. *Creating music which: uses a selection of composite (uses more than one exciter) rhythmic patterns; uses a variety of exciters; uses the exciters in a variety of ways; uses varying tempos and dynamics; and, has a sense of structure.*

2. *Creating music that: uses a selection of composite rhythmic patterns with some variation of tempo and some variety of ways of using the exciters.*

3. *Creating music that includes composite rhythmic patterns that involve more than one exciter or use more than one way of using the exciter(s).*

4. *Creating simple patterns that repeat using one exciter at a time.*

5. *Random picking up of objects with the intention to test rather than make music with occasional musical features.*

6. *Random picking up of objects with the intention to test rather than make music.*

This list could be expanded to show greater nuance. Drawing from the videos the examples below illustrate this spectrum of skill and musicality.

1 - Creating music which: uses a selection of composite (uses more than one exciter) rhythmic patterns; uses a variety of exciters; uses the exciters in a variety of ways; uses varying tempos and dynamics; and, has a sense of structure. *Child E slaps the green plastic leaf three times and then taps the paintbrush also three times. Then she scrapes the pickle stick quickly backwards and forwards finishing with two quick taps. She drops the pin pong ball then scrapes the wooden knife scrape in the same way as the pickle stick. This is followed by four slow taps of pickle stick, turning it over each time followed by fast scrape. The four slow taps are repeated on the wooden knife. NE interrupts unnecessarily and asks her to try some gentler ideas. She continues with slow then fast scraping of the metal object. She moves onto the black disc that she turns rhythmically before scraping it faster to finish.*

2 - Creating music that: uses a selection of composite rhythmic patterns with some variation of tempo and some variety of ways of using the exciters. *Child A picks up the pickle stick and wooden knife and says 'I'm thinking of what I'm going to do'. He slowly alternates between the two objects. Then he exchanges knife for ball. A clear repeating rhythmic pattern emerges - ball once, pickle stick twice. The pattern gets faster. This is followed by a slower section in which he uses the ball and pickle stick - sometimes alternating, sometimes together. Charlie puts the ball down and picks up the screw, continuing with the pickle stick in the other hand there is a further short slower meandering section. Exchanges screw for paintbrush - similar idea continues. Child I starts with eleven rapid taps using the wooden knife then rubs the coin quickly on the table. This is followed by two drops of a screw then she slides the chopstick on the table. She drops another coin, then the paper clip, and then creates two short bursts of rapid tapping using the pickle stick with a gap between each burst. Rapid tapping continues with the ruler. She drops the grey plastic bowl and does some scratching with the screw. Nancy asks her to think about finishing and she drops the grey plastic bowl once more to finish.*

3 - Creating music that includes composite rhythmic patterns that

involve more than one exciter or use more than one way of using the exciter(s). *Child B immediately picks up the pickle stick and the wooden knife, one in each hand. Using alternating hands he creates a pattern that goes 3-5 taps in one hand and then 3-5 taps in other hand. He exchanges the wooden knife for he paintbrush but keeps pickle stick. The same pattern continues. Then he exchanges the paintbrush for paper clip and carries on pattern but faster. Exchanges paper clip for screw and carries on pattern. Exchanges screw for ball. Sometimes taps ball sometimes drops it.*

4 - Creating simple patterns that repeat using one exciter at a time. *Child C makes a rhythmic pattern using the pickle stick that goes slow slow quick quick slow slow. She then repeats the rhythm using the metal object, then the black disc, then the grey plastic dish and then the wooden knife.*

5 - Random picking up of objects with the intention to test rather than make music with occasional musical features. *Child F starts with a few taps with the pickle stick, puts it down, then picks up the metal object and moves one part of it up and down. She abandons this to fiddle with the mic and ask what it's for. She drops the coin and then the ping pong ball sometimes catching it and sometimes letting t bounce. She drops two balls and they fall off the table. Slowly she taps the green plastic leaf then moves onto the paint brush creating a quick tap tap gap tap tap pattern. She picks up the wooden knife in her other hand and repeats the previous rhythm with both exciters. She finishes by picking up the pickle stick and flicking the wired spiral of it onto the table and dropping a marble. There is no clear ending and Nancy asks her to stop.*

6 - Random picking up of objects with the intention to test rather than make music *Child D lifts the knife and drops it, then twists the pickle stick then drops the black disc, followed by the ball. He continues by picking up and dropping different objects on after the other.*

B.3.5 How did the children learn to play Mogees?

It has already been mentioned, as with any instrument, Mogees needs to be practiced. Children also need to learn generic music skills to allow them to use any technology/software in practical music making - composing, improvising and performing [Savage and Butcher, 2007]. They were three different aspects of children's learning that could be observed in this study:

1 - Learning how to play Mogees technically as an instrument *This might mean: being able to change the sound reliably; discovering the sound affordances and techniques of the different exciters; being able to access them immediately when needed; being able to access the full range of notes available; being able to vary sounds; being able to use one hand to do one thing and the other hand to do something else; being able to recreate musical ideas from one occasion to another; and knowing how to use the capture function on what object to best or intended effect. As mentioned previously children cannot learn to play Mogees using the visual or architectural clues that are often available to them on conventional instruments. It requires a different approach. One such way would be inviting the children to set up the table with exactly which exciters they would like to be on it and placed where they would like them to be.*

2 - Learning how to improvise musically with Mogees *This might mean using musical elements expressively (dynamics, tempo, timbre), using the different affordances of Mogees, having musical intention, creating musical patterns that repeat and vary, organizing ideas into a coherent structure. Using Mogees as beginners the children do not need to worry about pitch as the sounds are already mapped onto various scales and so whatever they play it will sound pleasing, in a western tonal music sense. This allows them to focus on other elements such as timbre, tempo and dynamics as they don't feel like they are making mistakes by playing 'wrong' notes.*

3 - Learning to improvise with a partner using Mogees *A third part of* *and for this the children need to learn the skills of listening and coordinating musical ideas with a partner; adding complementary musical ideas; and, responding to the musical ideas of a partner. Though all instruments have some gestural qualities, the obvious gestural element of playing Mogees allows for easier matching of music/movement ideas between improvising partners. These could be observed in some of the pair improvisations.*

B.3.6 What pedagogies or strategies for adult's to support children's learning with Mogees might we begin to identify?

With second and third aspects of learning mentioned traditional pedagogies for supporting young people's creative music making individually and with others are relevant: These include:

- *Modeling improvising and composing including thinking out loud about your processes*
- *Talking about music you hear the children create - labeling and describing what you hear with both music specific but also rich everyday vocabulary*
- *Giving specific feedback and asking effective questions*
- *Setting different kinds of tasks and creating different kinds of opportunities*

Whatever the technologies used it is still the teacher and their understanding of pedagogy that will determine the effectives of the teaching and learning. With or without technologies that above pedagogies for supporting creative music making are not well understood by generalist primary teachers delivering whole class music teaching. As far the first aspect of learning how to play Mogees technically requires further study. Some conventional pedagogies associated with learning to play an instrument will be useful but other specific to Mogees will need to be developed.

B.3.7 What might be the learning potential of Mogees for children's music education?

With all teaching, whether learning about vibrations in science or learning how to improvise in music, achieving the desired learning outcome(s) and using the best *strategy for achieving this is paramount to teachers. Guidance for the use of ICT in music learning and teaching from the TTA (Teacher Training Agency) outlines three key principles for music educators regarding the use of music technology which are cited by Savage. These are:*

- *Decisions about when, when not and how to use ICT in lessons [should be] based on whether the use of ICT supports good practice in teaching music. If it doesn't don't use it.*
- *When planning, make sure that the use of ICT in a particular lesson or scheme of work directly relates to the chosen teaching and learning objectives.*
- *ICT should allow the teacher and student to learn something that could not be achieved without it; or allow them to learn something more effectively and efficiently than could otherwise be done.*

This where the adults understanding of the learning potential, affordances and limitations of the technology is crucial. Mogees in some respects can be treated like any other instrument. Ultimately it is the skill of the teacher in framing activities that will decide whether learning outcomes are achieved, not the technology used.

This said, Mogees is simple to use, allows for complexity and creativity without the barriers of learning how, for example, to bow a violin or get a correct embouchure on a trumpet. However, like any musical instrument, Mogees can be played by an expert as well as a novice and has its own learning trajectory.

What Mogees does offer is a type of music making that involves the whole body and harnesses, particularly young children's, multimodal creative play. Something that becomes separated or is lost once children are overcome with the technical challenges of playing an instrument or music is separated in the curriculum from movement and dance.

B.3.8 Does Mogees sustain the engagement of the children?

Within the short session Mogees did sustain the attention of all the children. However it is impossible within this very short time to ascertain whether this would continue over a number of sessions. This key issue is raised by Oore and cited by Savage and Butcher: the need to get the balance, in the software, *between simplicity, which can offer accessibility and complexity, which makes it less accessible, but sustains interest beyond the initial novelty. Further study of children's use of Mogees over time would be needed to see whether it was successful in this.*

B.3.9 What level of support is required for the participants to use the interface?

None of the children had any problems using the interface or were overwhelmed by the possibilities on offer. The only help needed was to find the second screen where the tuning function was. It is likely that once shown the children would have no problem finding this again on a subsequent occasion.

B.3.10 Were there any technical difficulties using the software?

There was the occasional need to turn the volume up and down depending on the app sound and when feedback occurred. There was occasional feedback The capture function produced variable results Children wanted to use the Capture function as if it was a traditional recording device: pressing once to start and once to stop. Pressing it a second just starts the process again. This confused the children and didn't help them produce the best result from this function. Though most children's focus was entirely on the music making a very few children seemed distracted and to watch very careful the flashing of the iPod which indicated sounds being made. Occasionally the sound the object/exciter made on the table was louder than the Mogees sound Most children misunderstood the capture feature thinking it was capturing the sound of any object not its unique vibrations, which could change depending on how it was played. Maybe a visual representation on the interface could be developed to support their understanding. Mogees was sometimes very sensitive to movement/vibrations on the floor and from voices.

B.3.11 Were there any technical difficulties using the hardware?

The length of lead of the contact microphone could be longer so as to be able to have the iPod more comfortable on a different surface from the contact microphone. Occasionally the contact microphone did not stick to the chosen object - usually a curved surface.

B.3.12 Any other notable interactions/activity?

The children's focus was almost entirely on the table, the exciters and therefore on making music. They spent very little time looking at the iPod. Most of the children only looked at the app when explicitly asked to. This is a welcome change to other available music technology.

Appendix C

Appendix: Interview with the composer Ed Handley

Below is reported an interview with Ed Handley about our one-year collaboration using Mogees, as described in section 6.

B: What are the main motivations that bring you to use Mogees ?

E: *I've been interested in physical modelling since I first read about it, I suppose about 10 years ago, when companies like Yamaha first released these instruments based on the Karplus-Strong thing, which was pretty basic and pretty intensive and expensive when it first came out. So initially it was the attraction to physical modelling, more of a sound generation thing than necessarily as a performance instrument. I think the instrument part of it became exciting once I'd seen it in action, when I'd seen the video. But more actually when I had the Mogees in front of me and I could actually have a play with it. I think that's when I could see the potential of it as a musical instrument. It has got quite a distinct sound compared to other physical modelling systems I've played with and it seems to be super-efficient. I've been playing around with* Modalys (audio physical modelling engine developed at IRCAM [mod], ndr) *for a few years and it's quite heavy on the CPU. And it's quite difficult to incorporate it into a live setting; it's fine doing pre-recordings or things like that or doing very minimal things with it, but to actually have it generating in a live situation was really difficult; I think with faster CPUs and things you can have a few instances now and get away with it, but the engine inside Mogees is much faster, and as I said,* *it's got its own sound. It is quite hard to put your finger on it. It doesn't work in the same way as the conventional physical models, as in you're not trying to necessarily replicate a string vibrating, you're not trying to replicate acoustic instruments, but it has a very acoustic sound and a very detailed sound, and I think the basic attraction of physical models obviously is how you excite them, so you've got this two sides to it; you've got the sound that's actually generating, but how you stimulate the model is just as important, and I think for me that's why it was attractive as a type of synthesis, because a lot of, you know, traditional ways of synthesising, you're taking a sine wave and then you're just manipulating it through other methods of modulation. With Mogees, the modulation is physical. The modulation is something very distinct, in that it's the exciter, it is what you chuck into the sound. So yes, it's interesting from all those points of view... and as an instrument, obviously I'm not a player, so from a professional point of view, I wasn't looking at it as an instrument for me to play, necessarily, but I think having used it, I realised actually I probably could play it. I can play, I'm just not very experienced at it in a live situation.*

B: Why do you think you could actually play the system, now that you tried it ?

E: *Because it doesn't require too much learning, as in really, you just need a sense of rhythm and a good object to play, whereas with most other instruments there are certain technicalities that you have to learn; obviously a drum, not so much, but even with a drum, obviously there's a certain amount of techniques... but say, pianos and strings and stuff... they take a long time to learn.*

B: Are you talking about the learning curve?

E: *Learning curve (with Mogees) is very short. You're just bashing an object, or scraping an object, so that's why when I saw it... it's something I could probably play, obviously with a little bit of practise to get the timing right, and the fact that it's... it isn't just actually tapping, it's also scraping, it's a whole load of gestures that can go into this thing that it can interpret.*

B: How your approach with the system changes between studio and live performances ?

E: *In regards to Mogees, obviously, there's only really been three occasions where we've used it in a live situation together...both had their own problems... but I think there were bound to be problems... I think the difference on how we've* *used Mogees on recorded material is obviously that it's a controlled environment. But the technique is the same, the main difference being that in the studio I may well excite the Mogees with another sample, with another piece of recorded sound... I may not actually play the Mogees conventionally, I would just use the model aspect of it... Just for control and timing... as I said, I'm not a player, so it makes sense if there's the option of pre-recording the excitation material, then I probably would, because then I can set it within a song structure a lot easier, because if I was to play, it would take me quite a few takes to get the timing right; but that's not such a huge distinction really, because in theory, you could do the same thing live if you wanted to be slightly perverse about it, actually play samples into the Mogees... but obviously one of the attractions of Mogees is that it's direct and you're using physical objects and obviously live you have an audience, and the audience often likes to see something happening, and they like to see something physical happening, so to have an instrument that you're playing is appropriate in a live situation and it's kind of more exciting for everyone. There's the option for better improvisation and a more interesting piece of music potentially.*

B: So you think that at the visual level, Mogees has some potential? How is it different from a classic acoustic or electronic instrument?

E: *Well, if you're comparing it to say a guitar... a guitar is a guitar, and it will be from the start of the set to an end of the set; there's a lot you can do with a guitar and there's various techniques for playing it; the difference with Mogees is that any object can effectively become a kind of instrument, which means throughout the performance you can utilise a whole different number of instruments, which obviously visually quite exciting for anyone that's watching , but also audibly you're getting this very diverse range of sounds and sound sources, so it has a lot of benefits... obviously the limitations of Mogees are compared to a regular instrument is the range of say notes you can access at any one time, you don't have complete freedom with that, and potentially the replication of a piece, can you play the same thing twice accurately. I haven't played with the latest version yet, and that might be a consideration. But if you're going for a virtuoso performance, obviously a lot of that is to be able to play the same thing almost exactly the same. And I'm sure probably it can do that now, I don't know. But the advantages outweigh the disadvantages, in that you've got this huge variety of sources.*

B: Now a more artistic question: what is the function of the sound of Mogees in your pieces ?

E: *I think because it has this acoustic quality and detail, especially when you increase the number of partials, you get this very intricate sound. And obviously you can hear the source in it: if it's a piece of wood, you can hear wood; you can hear the grain in the wood; you can hear the texture of the object that you're using. I think that's what distinguishes it from anything else, even from other physical models, because it has this particular way of doing things. And the buffer mode* (capture function, ndr) *is very interesting, in that it takes this frame of sound from a buffer, which also has a very distinct sound, you get these lovely harmonics. So I'd say it has such a distinct sound and that's why I would use it, because I can't think of anything else that really sounds like it.*

B: Are you encouraged to explore different types of physical, real-world objects through Mogees? if so, what are the main motivations ?

E: *Yeah, since I first saw Mogees... you always just to experiment and try different things with it.*

B: And why ?

E: *I think if you're interested in sound, I think you want to hear the object, you want to see how it propagates sound. How it vibrates. You just want to see how it sounding through Mogees. But I think that as a musician or a sound designer you're naturally inquisitive with these things, and even without the Mogees you'll be thinking about the acoustics of things and what happens if you tap this, what kind of sounds you get out of it. With the Mogees, it adds an extra dimension to that: you're getting this fusion out of whatever model you've chosen or created and the object.*

B: Would you use it for future compositions and live performances and if so, would you change anything ?

E: *I think definitely; I think there is probably room for a version of Mogees that is designed more for say the professional musician, potentially. Not because I think there is limitations with an app version, it's more because interfacing or getting an iPhone to communicate with other devices, say laptops and things like that and synchronising them.. It's possible, but it's not necessarily that convenient. And also a multi-channel version would be useful, from a compositional* *point of view, to be able to layer, without having to record and then record. To actually have multiple sound sources, all happening at once, would be great. To be able to plug in a few microphones and have them all going at once. I think from a musical experimentation that's interesting, because then you're hearing how these things work together all at once, as opposed to having to commit, recording one, and then trying something else. And also I suppose having a version with the full note ranges and perhaps a few ways of modulating some of the parameters of the model, for sure. I mean I have used the existed version of Mogees, or an existing MaxMSP version of Mogees, and it works really nicely and it's great fun to play with, but I can imagine a version that's just a little bit more tweaked for a classic kind of plug-in type scenario, where there are a few more functions than on the app version. But I think primarily the app version is the exciting thing, and I think the important thing, because I think that will attract everyone, not necessarily musicians.*

B: What do you think this technology can bring to musicians who could benefit from it ?

E: *I think any musician would enjoy playing it, and I think once you get over the wow' factor, because there is definitely a wow' factor, especially if you've not come across physical models...then I think once that initial excitement dies down, it can be used to make, you know, music and not... I think for improvising and playing it's a really great device, and obviously for a lot of musicians, that's key to what they do, certainly with composition, the play and the experimenting phase is the most important part, that's when you come up with the ideas. So anything that can enable a bit more freedom in that process, that's not a keyboard, or a string...that's unconventional, that encourages you to take, to do things differently. And that's the great thing about Mogees: it encourages you to look at the objects around you and think about them, or discover them in an acoustic sense, and there's not really been obviously a microphone, just a microphone is that to a degree, but that's not necessarily instantly musical whereas what the Mogees does is it takes these everyday sounds, these everyday objects, and it kinds of gives them a musicality instantly, and I think that's why. I think a lot of musicians would like it, because it encourages this experimentation, it encourages this discovery of the world around you.*

B: And do you think that this technology could be used for, say,

other domains of the human-computer interaction?

E: *Yeah...I think the whole idea of sonification'. There are so many fields that can apply to, and something like Mogees is a great way of sonifying an object, or a space. In terms of applications, there are a whole load of them, be they educational or therapeutic. I'm sure there's a lot, I'm not an expert in that field, so I don't know what, but I'd imagine that there's plenty of applications for it.*

Bibliography

Not for threes. URL http://en.wikipedia.org/wiki/Not_for_Threes. 150

Patchblocks. URL http://patchblocks.com/. 33

Cycling74. URL http://cycling74.com/products/max/. 36, 51

Mimio. URL http://www.mimio.com/en-EM.aspx. 40

Modalys. URL http://forumnet.ircam.fr/product/modalys/. 197

Plaid. URL http://www.plaid.co.uk/. 150

Praise project. URL http://www.iiia.csic.es/praise/. 134

Tai-chi project, 2009-2011. URL http://www.future-instruments.net/fr/taichi.php. 40

Science programmes of study: key stages 1 and 2. 9 2013. URL https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/239132/PRIMARY_national_curriculum_-_Science.pdf. 172

Jean-Marie Adrien. Representations of musical signals. In Giovanni De Poli, Aldo Piccialli, and Curtis Roads, editors, *Representations of musical signals*, chapter The missing link: modal synthesis, pages 269–298. MIT Press, Cambridge, MA, USA, 1991. ISBN 0-262-04113-8. 38

Roberto Mario Aimi. *Hybrid Percussion: Extending Physical Instruments Using Sampled Acoustics*. Massachusetts Institute of Technology, School of Architecture and Planning, Program in Media Arts and Sciences, 2007. 37

Newton Armstrong. *An Enactive Approach to Digital Musical Instrument Design*. Princeton University, 2006. 13

Federico Avanzini, Stefania Serafin, and Davide Rocchesso. Interactive simulation of rigid body interaction with friction-induced sound generation. *Speech and Audio Processing, IEEE Transactions on*, 13(5):1073–1081, 2005. 99

Ayah Bdeir and Ted Ullrich. Electronics as material: Littlebits. In *Proceedings of the Fifth International Conference on Tangible, Embedded, and Embodied Interaction*, TEI '11, pages 341–344, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-0478-8. doi: 10.1145/1935701.1935781. URL http://doi.acm.org/10.1145/1935701.1935781. 33

JD Beltran. telephonestory, 2001. URL http://identity.media.mit.edu/telephonestory.html. 39

Edgar Berdahl and Julius O. Smith. A tangible virtual vibrating string : A physically motivated virtual musical instrument interface. In Antonio Camurri, Gualtiero Volpe, and Stefania Serafin, editors, *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 299–302, Genoa, Italy, 2008. URL http://www.nime.org/proceedings/2008/nime2008_299.pdf. 38

Frédéric Bevilacqua, Remy Muller, and Norbert Schnell. MnM: a Max/MSP mapping toolbox. In *Proceedings of the 2005 conference on New interfaces for musical expression*, Vancouver, Canada, 2005. 51

Frédéric Bevilacqua, Nicolas Rasamimanana, Emmanuel Fléty, Serge Lemouton, and Florence Baschet. The augmented violin project: research, composition and performance report. In *Proceedings of the 2006 conference on New interfaces for musical expression*, NIME '06, pages 402–406, Paris, France, France, 2006. IRCAM Centre Pompidou. ISBN 2-84426-314-3. 16

Frédéric Bevilacqua, Fabrice Guédy, Norbert Schnell, Emmanuel Fléty, and Nicolas Leroy. Wireless sensor interface and gesture-follower for music pedagogy. In *Proceedings of the 7th international conference on New interfaces for musical expression*, NIME '07, pages 124–129, New York, NY, USA, 2007. ACM. doi: 10.1145/1279740.1279762. 6, 28, 48, 50, 51

Frédéric Bevilacqua, Bruno Zamborlin, Anthony Sypniewski, Norbert Schnell, Fabrice Guédy, and Nicolas Rasamimanana. Continuous realtime gesture following and recognition. In *Proceedings of the 8th international conference on Gesture in Embodied Communication and Human-Computer Interaction*, GW'09, pages 73–84, Berlin, Heidelberg, 2010. Springer-Verlag. ISBN 3-642-12552-2, 978-3-642-12552-2. doi: 10.1007/978-3-642-12553-9_7. 6, 28, 48, 50, 54, 70

Frédéric Bevilacqua, Florence Baschet, and Serge Lemouton. The augmented string quartet: Experiments and gesture following. *Journal of New Music Research*, 41(1):103–119, 2012. 28, 71

John Brooke. Sus-a quick and dirty usability scale. *Usability evaluation in industry*, 189:194, 1996. 120

Judith C Brown. Calculation of a constant q spectral transform. *The Journal of the Acoustical Society of America*, 89:425, 1991. 104

Baptiste Caramiaux, Jules Françoise, Norbert Schnell, and Frédéric Bevilacqua. Mapping through listening. *Computer Music Journal*, 38:34–48, 2014a. 26

Baptiste Caramiaux, Nicola Montecchio, Atau Tanaka, and Frédéric Bevilacqua. Adaptive gesture recognition with variation estimation for interactive systems. *ACM Transactions on Interactive Intelligent Systems*, 4(4):18:1–18:34, December 2014b. ISSN 2160-6455. doi: 10.1145/2643204. URL `http://doi.acm.org/10.1145/2643204`. 28

Jennie Carroll. Completing design in use: closing the appropriation cycle. In Timo Leino, Timo Saarinen, and Stefan Klein, editors, *ECIS*, pages 337–347, 2004. 18, 23, 130, 147

Roudra Chakraborty, Debapriya Sengupta, and Sagnik Sinha. Pitch tracking of acoustic signals based on average squared mean difference function. *Signal, Image and Video Processing*, 3(4):319–327, 2009. 114

Nisha Checka. *A System for Tracking and Characterizing Acoustic Impacts on Large Interactive Surfaces*. Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science, 2001. 39

Miha Ciglar. An ultrasound based instrument generating audible and tactile sound. In *proc. NIME*, 2010. 42

Gabe Cohn, Dan Morris, and SN Patel. Your noise is my command: sensing gestures using the body as an antenna. *Proceedings of the CHI 2011*, pages 791–800, 2011. 91

Gabe Cohn, Daniel Morris, Shwetak Patel, and Desney Tan. Humantenna: using the body as an antenna for real-time whole-body interaction. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems*, CHI '12, pages 1901–1910, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1015-4. doi: 10.1145/2208276.2208330. 91

Beginner's Mind Collective and David Shaw. Makey makey: Improvising tangible and nature-based user interfaces. In *Proceedings of the Sixth International Conference on Tangible, Embedded and Embodied Interaction*, TEI '12, pages 367–370, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1174-8. doi: 10.1145/2148131.2148219. URL http://doi.acm.org/10.1145/2148131.2148219. 33

Andrew A. Cook and Graham Pullin. Tactophonics: your favourite thing wants to sing. In *Proceedings of the 7th international conference on New interfaces for musical expression*, NIME '07, pages 285–288, New York, NY, USA, 2007. ACM. doi: 10.1145/1279740.1279801. 35, 36, 91, 127

Perry R Cook. Toward physically-informed parametric synthesis of sound effects. In *Applications of Signal Processing to Audio and Acoustics, 1999 IEEE Workshop on*, pages 1–5. IEEE, 1999. 99

Perry R. Cook. Principles for designing computer music controllers. In Ivan Poupyrev, Michael J. Lyons, Sidney S. Fels, and Tina Blaine, editors, *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 3–6, Seattle, WA, 2001. URL http://www.nime.org/proceedings/2001/nime2001_003.pdf. 2, 17

Alain Crevoisier and Cédric Bornand. Transforming daily life objects into tactile interfaces. In *Proceedings of the 3rd European Conference on Smart Sensing* *and Context*, EuroSSC '08, pages 1–13, Berlin, Heidelberg, 2008a. Springer-Verlag. ISBN 978-3-540-88792-8. doi: 10.1007/978-3-540-88793-5_1. 40, 41, 42

Alain Crevoisier and Cdric Bornand. Transforming daily life objects into tactile interfaces. In *Smart Sensing and Context*, volume 5279 of *Lecture Notes in Computer Science*, pages 1–13. Springer Berlin Heidelberg, 2008b. URL `http://dx.doi.org/10.1007/978-3-540-88793-5_1`. 30

Alain Crevoisier and Pietro Polotti. Tangible acoustic interfaces and their applications for the design of new musical instruments. *New Interfaces for Musical Expression (NIME 2005)*, pages 97–100, 2005. 30, 37

Mihaly Csikszentmihalyi. *Flow: The psychology of optimal experience*, volume 41. HarperPerennial New York, 1991. 13

Alain de Cheveigné and Hideki Kawahara. Yin, a fundamental frequency estimator for speech and music. *J Acoust Soc Am*, 111:1917–1930, 2002. 114

Alan Dix. Designing for appropriation. In *Proceedings of the 21st British HCI Group Annual Conference on People and Computers: HCI...but not as we know it - Volume 2*, BCS-HCI '07, pages 27–30, Swinton, UK, UK, 2007. British Computer Society. ISBN 978-1-902505-95-4. 3, 19

Paul Dourish. *Where The Action Is: The Foundations Of Embodied Interaction*. Bradford books. Bradford Books, 2004. ISBN 9780262541787. 4, 11, 18

Jeff Dyck, David Pinelle, Barry Brown, and Carl Gutwin. Learning from games: Hci design innovations in entertainment software. In *Graphics Interface*, pages 237–246. Citeseer, 2003. 3

Gerhard Eckel, Francisco Iovino, and René Caussé. Sound synthesis by physical modelling with modalys. In *Proc. International Symposium on Musical Acoustics*, pages 479–482, 1995. 111

Georg Essl and Sile O'Modhrain. An enactive approach to the design of new tangible musical instruments. *Organised Sound*, 11(03):285, November 2006. ISSN 1355-7718. doi: 10.1017/S135577180600152X. 13, 31

Jerry Fails and Dan Olsen. A design tool for camera-based interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '03, pages 449–456, New York, NY, USA, 2003a. ACM. ISBN 1-58113-630-7. doi: 10.1145/642611.642690. 25

Jerry Alan Fails and Dan R. Olsen, Jr. Interactive machine learning. In *Proceedings of the 8th international conference on Intelligent user interfaces*, IUI '03, pages 39–45, New York, NY, USA, 2003b. ACM. ISBN 1-58113-586-6. doi: 10.1145/604045.604056. 25

Andy Farnell and Koray Tahiroglu. Helsinki new media lab, 20th anniversary seminar. 2014. URL http://goo.gl/CVZ2WE. 171

Rebecca Fiebrink. Real-time interaction with supervised learning. In *CHI '10 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '10, pages 2935–2938, New York, NY, USA, 2010. ACM. ISBN 978-1-60558-930-5. doi: 10.1145/1753846.1753889. 85

Rebecca Fiebrink, Perry R. Cook, and Dan Trueman. Human model evaluation in interactive supervised learning. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pages 147–156, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-0228-9. doi: 10.1145/1978942.1978965. 26, 57, 67, 85

Mathias Fink and Claire Prada. Acoustic time-reversal mirrors. *Inverse Problems*, 17(1):R1, 2001. 40

Masaki Fujihata, Kiyoshi Furukawa, and W Münch. Notes on Small Fish. *Ars Electronica 2000 Festival ...*, 2000. 32

Susan Gasson. Human-centered vs. user-centered approaches to information system design. *Journal of Information Technology Theory and Application (JITTA)*, 5(2):29–46, 2003. 20

Emilien Ghomi, Guillaume Faure, Stéphane Huot, Olivier Chapuis, and Michel Beaudouin-Lafon. Using rhythmic patterns as an input method. *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems - CHI '12*, page 1253, 2012. doi: 10.1145/2207676.2208579. 43

F. Gmeiner. Table recorder, 2007. URL http://www.youtube.com/watch?v=ilNj4gXGfAk. 35

Michael Gurevich and Andrew Fyans. Digital musical interactions: Performersystem relationships and their perception by spectators. *Organised Sound*, pages 166–175, 2011. 24

Michael Gurevich, Adnan Marquez-Borbon, and Paul Stapleton. Playing with constraints: Stylistic variation with a simple electronic instrument. *Computer Music Journal*, 36(1):23–41, 2012. doi: 10.1162/COMJ_a_00103. 21

Kjetil Falkenberg Hansen. *The turntable: The instrument of hip-hop*, page Forthcoming. Cambridge University Press, 2014. 15

Chris Harrison and Scott E. Hudson. Scratch input: creating large, inexpensive, unpowered and mobile finger input surfaces. In *Proceedings of the 21st annual ACM symposium on User interface software and technology*, UIST '08, pages 205–208, New York, NY, USA, 2008. ACM. ISBN 978-1-59593-975-3. doi: 10.1145/1449715.1449747. 42, 43

Chris Harrison, Desney Tan, and Dan Morris. Skinput: appropriating the body as an input surface. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '10, pages 453–462, New York, NY, USA, 2010. ACM. ISBN 978-1-60558-929-9. doi: 10.1145/1753326.1753394. 43, 91

Chris Harrison, Hrvoje Benko, and Andrew D Wilson. OmniTouch : Wearable Multitouch Interaction Everywhere. In *Proceedings of the 24th annual ACM symposium on User interface software and technology*, pages 441–450, 2011a. ISBN 9781450307161. 91

Chris Harrison, Julia Schwarz, and Scott E. Hudson. Tapsense: enhancing finger interaction on touch surfaces. In *Proceedings of the 24th annual ACM symposium on User interface software and technology*, UIST '11, pages 627–636, New York, NY, USA, 2011b. ACM. ISBN 978-1-4503-0716-1. doi: 10.1145/2047196.2047279. 42

Chris Harrison, Robert Xiao, and Scott E. Hudson. Acoustic barcodes: passive, durable and inexpensive notched identification tags. In *UIST*, pages 563–568, 2012. 43

Björn Hartmann, Leith Abdulla, Manas Mittal, and Scott R. Klemmer. Authoring sensor-based interactions by demonstration with direct manipulation and pattern recognition. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '07, pages 145–154, New York, NY, USA, 2007. ACM. ISBN 978-1-59593-593-9. doi: 10.1145/1240624.1240646. 25

Ian Hattwick, Preston Beebe, Zachary Hale, Marcelo Wanderley, Phillippe Leroux, and Fabrice Marandola. Unsounding objects: Audio feature extraction for the control of sound synthesis. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, 2014. 38

M. Hauenstein. Mark hauenstein's website, 2004. URL `http://www.nurons.net/audioshaker/about.htm`. 33

Simon Haykin. Array signal processing. *Englewood Cliffs, NJ, Prentice-Hall, Inc., 1985, 493 p. For individual items see A85-43961 to A85-43963.*, 1, 1985. 39

RK Ing, S. Catheline, N. Quieffin, and M. Fink. Dynamic focusing using a unique transducer and time reversal process. In *Proceeding of the 8th International Congress on Sound and Vibration*, pages 119–126, 2001. 40

Hiroshi Ishii, Craig Wisneski, Julian Orbanes, Ben Chun, and Joe Paradiso. Pingpongplus: design of an athletic-tangible interface for computer-supported cooperative play. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, CHI '99, pages 394–401, New York, NY, USA, 1999. ACM. ISBN 0-201-48559-1. doi: 10.1145/302979.303115. 39

Hiroshi Ishii, Ali Mazalek, and Jay Lee. Bottles as a minimal interface to access digital information. In *CHI '01 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '01, pages 187–188, New York, NY, USA, 2001. ACM. ISBN 1-58113-340-5. doi: 10.1145/634067.634180. 33

Robert J.K. 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*, CHI '08, pages 201–210, New York, NY, USA, 2008. ACM. ISBN 978-1-60558-011-1. doi: 10.1145/1357054.1357089. 30, 91

D. A. Jaffe and J. O. Smith. Extensions of the Karplus-Strong plucked-string algorithm. *Computer Music J.*, 7(2):56–69, 1983. 95

Alexander Refsum Jensenius. To gesture or not? An analysis of terminology in NIME proceedings 2001–2013. In *Proceedings of the International Conference on New Interfaces For Musical Expression*, pages 217–220, London, 2014. 24

Alexander Refsum Jensenius, Marcelo M. Wanderley, Rolf Inge Godøy, and Marc Leman. Musical gestures: Concepts and methods in research. In Rolf Inge Godøy and Marc Leman, editors, *Musical gestures: Sound, movement, and meaning*, pages 12–35. Routledge, New York, 2010. 24

Sergi Jordà. Digital instruments and players part ii: Diversity, freedom and control. In *Proceedings of the International Computer Music Conference*, pages 706–710, 2004a. 1, 17, 129, 171

Sergi Jordà. Digital instruments and players: part i—efficiency and apprenticeship. In *Proceedings of the 2004 conference on New interfaces for musical expression*, pages 59–63. National University of Singapore, 2004b. 21

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*, TEI '07, pages 139–146, New York, NY, USA, 2007. ACM. ISBN 978-1-59593-619-6. doi: 10.1145/1226969.1226998. 32

Kevin Karplus and Alex Strong. Digital Synthesis of Plucked String and Drum Timbres. *Computer Music Journal*, 7(2):43–55, 1983. 95

Sven Kratz and Rafael Ballagas. Unravelling seams: improving mobile gesture recognition with visual feedback techniques. In *Proceedings of the SIGCHI*

Conference on Human Factors in Computing Systems, CHI '09, pages 937–940, New York, NY, USA, 2009. ACM. ISBN 978-1-60558-246-7. doi: 10.1145/1518701.1518844. 79

Marc Leman. *Embodied Music Cognition and Mediation Technology.* MIT Press, 2008. ISBN 9780262122931. 13, 30

J. Linjama, P. Korpipää, Juha Kela, and T. Rantakokko. ActionCube: a tangible mobile gesture interaction tutorial. In *Proceedings of the 2nd international conference on Tangible and embedded interaction*, pages 169–172. ACM, 2008. ISBN 9781605580043. 79

D. Brandon Lloyd, Nikunj Raghuvanshi, and Naga K. Govindaraju. Sound synthesis for impact sounds in video games. In *Symposium on Interactive 3D Graphics and Games*, I3D '11, pages 55–62 PAGE@7, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-0565-5. doi: 10.1145/1944745.1944755. 102

Pedro Lopes, Ricardo Jota, and Joaquim A. Jorge. Augmenting touch interaction through acoustic sensing. *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces - ITS '11*, page 53, 2011. doi: 10.1145/2076354.2076364. 42

Hao Lü and Yang Li. Gesture coder: a tool for programming multi-touch gestures by demonstration. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems*, CHI '12, pages 2875–2884, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1015-4. doi: 10.1145/2208636.2208693. 26

Wendy E Mackay. Triggers and barriers to customizing software. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 153–160. ACM, 1991. 3

Thor Magnusson. *Epistemic Tools: The Phenomenology of Digital Musical Instruments.* PhD thesis, Brighton University, 2009. 3, 14

Steve Mann. Natural interfaces for musical expression: Physiphones and a physics-based organology. In *Proceedings of the 7th international conference on New interfaces for musical expression*, pages 118–123. ACM, 2007. 36

Sampada Marathe and S Shyam Sundar. What drives customization: control or identity? In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 781–790. ACM, 2011. 3

James McDermott, Toby Gifford, Anders Bouwer, and Mark Wagy. *Should Music Interaction Be Easy?*, chapter 2, pages 29–47. Springer, London, 2013. ISBN 978-1-4471-2989-9. 21

Andrew McPherson and Youngmoo E. Kim. The problem of the second performer: Building a community around an augmented piano. *Computer Music Journal*, 36(4):10–27, 2012. 16

Andrew McPherson, Adrian Gierakowski, and Adam M. Stark. The space between the notes: Adding expressive pitch control to the piano keyboard. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, pages 2195–2204, New York, NY, USA, 2013. ACM. ISBN 978-1-4503-1899-0. doi: 10.1145/2470654.2481302. 16

Richard Menary. *The extended mind*. Bradford Books. Mit Press, 2010. ISBN 9780262014038. 12

David Merrill and Hayes Raffle. The sound of touch. *CHI '07 extended abstracts on Human factors in computing systems - CHI '07*, page 2807, 2007. doi: 10.1145/1240866.1241044. 37

David Merrill, Hayes Raffle, and Roberto Aimi. The sound of touch: physical manipulation of digital sound. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '08, pages 739–742, New York, NY, USA, 2008. ACM. ISBN 978-1-60558-011-1. doi: 10.1145/1357054.1357171. 37

Sushmita Mitra and Tinku Acharya. Gesture Recognition: A Survey. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 37(3):311–324, 2007. doi: 10.1109/TSMCC.2007.893280. 27

Stefano Delle Monache, Pietro Polotti, and Davide Rocchesso. A toolkit for explorations in sonic interaction design. In *Proceedings of the 5th Audio Mostly Conference: A Conference on Interaction with Sound*, AM '10, pages 1:1–1:7, New York, NY, USA, 2010. ACM. ISBN 978-1-4503-0046-9. doi: 10.1145/1859799.1859800. 36, 37

F. Richard Moore. The dysfunctions of MIDI. *Computer music journal*, 1988. 13

Dan Newton and Mark T Marshall. The Augmentalist : Enabling Musicians to Develop Augmented Musical Instruments. *Proceedings of the fifth international conference on Tangible, Embedded, and Embodied Interaction*, pages 249–252, 2011. 2, 13, 16, 171

Michael Nielsen, Moritz Strring, Thomas B. Moeslund, and Erik Granum. A procedure for developing intuitive and ergonomic gesture interfaces for man-machine interaction. In *Gesture-Based Communication in Human-Computer Interaction Lecture Notes in Computer Science. Volume 2915*, pages 409–420, 2004. 53

M. Niinimaki. Mixmaster project's website, 2009. URL http://originalhamsters.com/motion/mixedup.php. 33

Donald A. Norman. *The Invisible Computer*. MIT Press, Cambridge, MA, USA, 1998. ISBN 0-262-14065-9. 24, 53, 91

Sile O'Modhrain and Georg Essl. Pebblebox and crumblebag: tactile interfaces for granular synthesis. In *Proceedings of the 2004 conference on New interfaces for musical expression*, NIME '04, pages 74–79, Singapore, Singapore, 2004. National University of Singapore. 35, 91

Nicola Orio, Norbert Schnell, and M.M. Wanderley. Input devices for musical expression: borrowing tools from HCI. In *Proceedings of the 2001 conference on New interfaces for musical expression*, pages 1–4. National University of Singapore, 2001. 166

Stanley R Page, Todd J Johnsgard, Uhl Albert, and C Dennis Allen. User customization of a word processor. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 340–346. ACM, 1996. 3

J.a. Paradiso and N. Checka. Passive acoustic sensing for tracking knocks atop large interactive displays. *Proceedings of IEEE Sensors*, pages 521–527, 2002. doi: 10.1109/ICSENS.2002.1037150. 39

Joseph Paradiso and Kai Yuh Hsiao. Swept-frequency, magnetically-coupled resonant tags for realtime, continuous, multiparameter control. In *CHI '99 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '99, pages 212–213, New York, NY, USA, 1999. ACM. ISBN 1-58113-158-5. doi: 10.1145/632716.632848. 33

Joseph a. Paradiso and Che King Leo. Tracking and characterizing knocks atop large interactive displays. *Sensor Review*, 25(2):134–143, 2005. ISSN 0260-2288. doi: 10.1108/02602280510585727. 39

Joseph A Paradiso, Che King Leo, Nisha Checka, and Kaijen Hsiao. Passive Acoustic Knock Tracking for Interactive Windows. *October*, 2002. 39

James Patten. Audiopad: A tag-based interface for musical performance. In *Proceedings of the 2002 conference on New Interfaces for Musical Expression*, pages 11–16, 2002. 32

D. T. Pham, Ze Ji, Ming Yang, Zuobin Wang, and Mostafa Al-Kutubi. A novel human-computer interface based on passive acoustic localisation. In *Proceedings of the 12th international conference on Human-computer interaction: interaction platforms and techniques*, HCI'07, pages 901–909, Berlin, Heidelberg, 2007. Springer-Verlag. ISBN 978-3-540-73106-1. 40, 41

Plaid. Reachy prints. Warp records, 2014. 9

Pietro Polotti and Stefano Delle Monache. Gamelunch: forging a dining experience through sound. *CHI'08 extended ...*, pages 1–6, 2008. 38

Pietro Polotti, Manuel Sampietro, Augusto Sarti, Stefano Tubaro, and Alain Crevoisier. Acoustic localization of tactile interactions for the development of novel tangible interfaces. In *Proc. of the 8th Int. Conference on Digital Audio Effects (DAFX-05)*, 2005a. 39, 40

Pietro Polotti, Manuel Sampietro, Augusto Sarti, Stefano Tubaro, and Alain Crevoisier. Acoustic localization of tactile interactions for the development of novel tangible interfaces. In *Proc. of the 8th Int. Conference on Digital Audio Effects (DAFX-05)*, 2005b. 30

Otniel Portillo-Rodrguez, Oscar Osvaldo Sandoval-Gonzalez, Emanuele Ruffaldi, Rosario Leonardi, Carlo Alberto Avizzano, and Massimo Bergamasco. Real-time gesture recognition, evaluation and feed-forward correction of a multimodal tai-chi platform. In Antti Pirhonen and Stephen A. Brewster, editors, *HAID*, volume 5270 of *Lecture Notes in Computer Science*, pages 30–39. Springer, 2008. ISBN 978-3-540-87882-7. 27

Nissanka B. Priyantha, Anit Chakraborty, and Hari Balakrishnan. The cricket location-support system. In *Proceedings of the 6th annual international conference on Mobile computing and networking*, pages 32–43. ACM, 2000. 41

Miller Puckette. Infuriating nonlinear reverberator. In *proc. ICMC*, 2011. 34, 38

MS Puckette, MSP Ucsd, and Theodore Apel. Real-time audio analysis tools for Pd and MSP analysis. *International Computer Music Conference*, pages 109–112, 1998. 42, 104

Lawrence R. Rabiner. A tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. *Proceedings of the IEEE*, 77(2):257–286, 1989a. ISSN 0018-9219. 177

Lawrence R Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition. In *Proceedings of the IEEE*, pages 257–286, 1989b. doi: 10.1109/5.18626. 57

Nicolas Rasamimanana, Norbert Schnell, Emmanuel Flety, Bruno Zamborlin, Ircam Cnrs Stms, Igor Stavinsky, and Paris France. Modular Musical Objects Towards Embodied Control Of Digital Music Real Time Musical Interactions. *Design*, pages 1–4, 2011. 33, 36, 91

Zhimin Ren, Hengchin Yeh, and Ming C Lin. Example-guided physically based modal sound synthesis. *ACM Transactions on Graphics (TOG)*, 32(1):1, 2013. 102

John Richards. Beyond diy in electronic music. *Organised Sound*, pages 274–281, 2013. 166

H.J. Rindorf. *Acoustic Emission Source Location in Theory and in Practice*. Technical review. Brüel & Kjær, 1981. 39

Davide Rocchesso and Federico Fontana. *The sounding object*. Mondo estremo, 2003. 37, 99

Jaime Ruiz, Yang Li, and Edward Lank. User-defined motion gestures for mobile interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pages 197–206, New York, NY, USA, 2011a. ACM. ISBN 978-1-4503-0228-9. doi: 10.1145/1978942.1978971. 26

Jaime Ruiz, Yang Li, and Edward Lank. User-defined motion gestures for mobile interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pages 197–206, New York, NY, USA, 2011b. ACM. ISBN 978-1-4503-0228-9. doi: 10.1145/1978942.1978971. 25

Jonathan Savage and Jason Butcher. Dubdubdub: Improvisation using the sounds of the world wide web. *Journal of Music, Technology & Education*, 1(1):83–96, 2007. 186, 192

Daniel Schlessinger and Julius O. Smith. The kalichord : A physically modeled electro-acoustic plucked string instrument. In Noel Zahler, Roger B. Dannenberg, and Tom Sullivan, editors, *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 98–101, Pittsburgh, PA, United States, 2009. URL `http://www.nime.org/proceedings/2009/nime2009_098.pdf`. 38

Norbert Schnell, Axel Röbel, Diemo Schwarz, Geoffroy Peeters, and Riccardo Borghesi. Mubu and friends: Assembling tools for content based real-time interactive audio processing in max/msp. In *Proceedings of the International Computer Music Conference (ICMC)*, 2009. 51, 111

Norbert Schnell, Fr'ed'eric Bevilacqua, Nicolas Rasamimanana, Julien Blois, Fabrice Gu'edy, and Emmanuel Fl'ety. Playing the “mo” ? gestural control and re-embodiment of recorded sound and music. In Alexander R. Jensenius, Anders Tveit, Rolf I. Godo y, and Dan Overholt, editors, *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 535–536, Oslo, Norway, 2011. URL `http://www.nime.org/proceedings/2011/nime2011_535.pdf`. 33, 36, 111

Diemo Schwarz. Corpus-based concatenative synthesis. *Signal Processing Magazine, IEEE*, 24(2):92–104, 2007. 37

Diemo Schwarz and Xavier Rodet. Spectral envelope estimation and representation for sound analysis-synthesis. In *proc. ICMC*, pages 351–354. Citeseer, 1999. 103

Diemo Schwarz, Pierre Alexandre Tremblay, and Alex Harker. Rich contacts: Corpus-based convolution of contact interaction sound for enhanced musical expression. In Baptiste Caramiaux, Koray Tahiroglu, Rebecca Fiebrink, and Atau Tanaka, editors, *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 247–250, London, United Kingdom, June 30 – July 03 2014. Goldsmiths, University of London. URL http://www.nime.org/proceedings/2014/nime2014_451.pdf. 37

Stefania Serafin. *The sound of friction: real-time models, playability and musical applications*. PhD thesis, stanford university, 2004. 99

Orit Shaer. Tangible User Interfaces: Past, Present, and Future Directions. *Foundations and Trends in HumanComputer Interaction*, (1-2):1–137, 2009. ISSN 1551-3955. doi: 10.1561/1100000026. 31, 32, 39

Julius O. Smith. Physical modeling using digital waveguides. *Computer Music Journal*, pages 1–28, 1992. 38

Janis Sokolovskis and Andrew P McPherson. Optical measurement of acoustic drum strike locations. In *proc. NIME*, 2014. 40

D. Stanley. The abstractmachine's website. URL http://www.abstractmachine.net/blog/abstractmachinev87d6/. 33

Italo Testa, Gianpaolo Evangelista, and Sergio Cavaliere. Physically inspired models for the synthesis of stiff strings with dispersive waveguides. *EURASIP Journal on Advances in Signal Processing*, 2004(7):964–977, 2004. 99

Pavan Turaga, Rama Chellappa, VS Subrahmanian, and Octavian Udrea. Machine recognition of human activities: A survey. *Circuits and Systems for Video Technology, IEEE Transactions on*, 18(11):1473–1488, 2008. ISSN 1051-8215. 27

Brygg Ullmer and Hiroshi Ishii. Emerging frameworks for tangible user interfaces. *IBM Syst. J.*, 39(3-4):915–931, July 2000. ISSN 0018-8670. doi: 10.1147/sj.393.0915. 31

Yon Visell and Jeremy Cooperstock. Enabling gestural interaction by means of tracking dynamical systems models and assistive feedback. In *Systems, Man and Cybernetics, 2007. ISIC. IEEE International Conference on*, pages 3373–3378, 2007. doi: 10.1109/ICSMC.2007.4414093. 27

Gil Weinberg and Seum-Lim Gan. The squeezables: Toward an expressive and interdependent multi-player musical instrument. *Comput. Music J.*, 25(2):37–45, July 2001. ISSN 0148-9267. doi: 10.1162/014892601750302570. 33

David Wessel and Matthew Wright. Problems and Prospects for Intimate Musical Control of Computers. *Computer Music Journal*, 26(3):11–22, September 2002. ISSN 0148-9267. doi: 10.1162/014892602320582945. 2, 13, 21, 22, 147

John Williamson. *Continuous uncertain interaction*. PhD thesis, University of Glasgow, 2006. 27

John Williamson. Stane : Synthesized Surfaces for Tactile Input. *Design*, 2008. 35

Andrew D. Wilson. Touchlight: An imaging touch screen and display for gesture-based interaction. In *In ICMI 04: Proceedings of the 6th international conference on Multimodal interfaces*, pages 69–76. ACM Press, 2004. 42

Andrew D. Wilson and Aaron F. Bobick. Parametric hidden markov models for gesture recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 21(9):884–900, 1999. 27

Jacob O. Wobbrock, Meredith Ringel Morris, and Andrew D. Wilson. User-defined gestures for surface computing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, pages 1083–1092, New York, NY, USA, 2009. ACM. ISBN 978-1-60558-246-7. doi: 10.1145/1518701.1518866. 25, 59

Robert Xiao, Chris Harrison, and Scott E. Hudson. Worldkit: Rapid and easy creation of ad-hoc interactive applications on everyday surfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, pages 879–888, New York, NY, USA, 2013. ACM. ISBN 978-1-4503-1899-0. doi: 10.1145/2470654.2466113. 91

Victor Zappi and Andrew McPherson. Dimensionality and appropriation in digital musical instrument design. In *Proc. New Interfaces for Musical Expression*, 2014. 21

Amit Zoran and Joseph A Paradiso. The chameleon guitar?guitar with a replaceable resonator. *Journal of New Music Research*, 40(1):59–74, 2011. 38