Digital Musical Instruments
for People with Physical Disabilities

Zacharias Vamvakousis
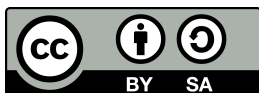
To the ones that remind us how beautiful life is...

Acknowledgements

First of all, I would like to express my deep gratitude to my supervisor Rafael Ramirez for giving me the opportunity to do this PhD and for supporting me in this long journey. He had the charisma to make things look clear in moments that I felt lost.

I am also grateful that all these years I had the opportunity to be part of Music Technology Group. It has been a really nice environment to do research, to discuss ideas and to make good friends. I was lucky for sharing the same office with Sergio Giraldo. We had the opportunity to not only discuss about research in music technology, but also play music together in our free time. He has been a great companion all these years and a good friend. I am also grateful for the nice moments we shared with all my good friends in MTG and the other research groups of in the campus.

I am very thankful to my good friend Kostas Manikis, whom I met when a friend of mine asked me if I could construct a Digital Musical Instrument for him. I am really happy with our cooperation and friendship. We implemented together his controller, as his ideas and feedback were decisive. I would also like to thank Kostas's friend, Dimitris Fotopoulos, for constructing the wooden body of the controller.

I would also like to express my gratitude to Joan Sandoval for asking me to supervise his Bachelor's Degree Final Project. Through him I met Laura de Pablos. We worked along with Joan to implement a guitar controller for Laura. Laura's feedback and ideas were very helpful. Interacting with her was always very inspiring.

Special thanks go to Marcel Schmith, for asking us to supervise him in his

Abstract

Playing a musical instrument has been shown to have a positive impact in the life of individuals in many different ways. Nevertheless, due to physical disabilities, some people are unable to play conventional musical instruments. In this dissertation, we consider different types of physical disabilities and implement specific digital musical instruments suitable for people with disabilities of each type. Firstly, we consider the case of people with limited sensorimotor upper limb functions, and we construct low-cost digital instruments for three different scenarios. Results indicate that the constructed prototypes allow musical expression and improve the quality of life of these users. Secondly, we consider disabilities such as tetraplegia or locked-in syndrome with unaffected eye-movements. For individuals with such conditions, we propose the EyeHarp, a gaze-controlled digital music instrument, and develop specific target selection algorithms which maximize the temporal and spatial accuracy required in music performance. We evaluate the instrument on subjects without physical disabilities, both from an audience and performer perspective. Results indicate that the EyeHarp has a steep learning curve and it allows expressive music performances. Finally, we examine the case of brain-controlled music interfaces. We mainly focus in auditory event related potential-based interfaces. In particular, we investigate and evaluate how timbre, pitch and spatialization auditory cues affect the performance of such interfaces.

Resumen

Se ha demostrado que tocar instrumentos musicales tiene un impacto positivo en muchos aspectos de la vida de las personas. Sin embargo, debido a discapacidades fsicas, a un gran nmero de personas les es imposible tocar instrumentos musicales tradicionales. En esta tesis doctoral consideramos diferentes tipos de discapacidades fsicas e implementamos instrumentos musicales digitales adaptados a las capacidades de las personas que las padecen. En primer lugar, consideramos el caso de personas con discapacidad motora en los miembros superiores sensoriomotoras. Utilizando materiales de bajo coste implementamos prototipos en tres escenarios diferentes. Los resultados indican que los prototipos construidos permiten la expresin musical y mejoran la calidad de vida de los usuarios. En segundo lugar, consideramos discapacidades como la tetraplejia o el sndrome locked-in donde aun se conservan los movimientos oculares. En ese caso, se propone el EyeHarp, un instrumento que se controla con movimientos de los ojos. Hemos desarrollado algoritmos de seleccin que maximizan la precisin temporal y espacial requerida en la ejecucin de instrumentos musicales y evaluamos el instrumento con gente sin discapacidades, desde la perspectiva de la audiencia y del msico. Los resultados indican que el EyeHarp tiene una curva de aprendizaje inclinada y permite interpretaciones musicales expresivas. Finalmente examinamos el caso de las interfaces musicales cerebro-ordenador. En particular, investigamos interfaces cerebro-ordenador basadas en potenciales relacionados con eventos auditivos. Investigamos cmo timbre, tono y espacializacin afectan el rendimiento de dichas interfaces y proponemos y evaluamos interfaces musicales basadas en esta tcnica.

Contents

List of Figures

List of Tables

Introduction

If I were not a physicist, I would probably be a musician. I often think in music. I live my daydreams in music. I see my life in terms of music... I cannot tell if I would have done any creative work of importance in music, but I do know that I get most joy in life out of my violin.

Albert Einstein

1.1 A personal Introduction to the thesis

Playing music has a very important role in my life. The idea of not being able to play music is stressful. My desire to make research on digital Musical Instruments (DMIs) designed for people with motor disabilities was born nine years ago, when a friend that plays the Cretan lira, an instrument similar to the violin, had a serious accident with his motorbike. The first two days it was not clear whether he would be able to move his upper limbs. Fortunately, although he is now paraplegic, he is still able to play lira, guitar and sing. I became even more sensitive about the quality of life of people with motor disabilities after watching a movie called *The Sea Inside* . This movie is about the true story of a man's 30-year battle for the legal right to end his own life. After watching that movie I realized that one important thing that was missing from this man's life was the opportunity to be creative. Having the opportunity to be creative and contribute to the society in some way would increase the desire to live. Music gives pleasure

to both the performer and the audience. As a result music benefits the performer in many ways. Most notably: (i) by expressing one's feelings and (ii) by making him feel useful to the others. Being a musician and a computer scientist I thought that I might contribute to the improvement of the quality of life of people with motor disabilities by inventing Digital Musical Instruments adapted to the special needs of each of them.

The biggest part of this thesis is about gaze and brain-controlled music interfaces. The reason for that is that this kind of interfaces can be controlled even by people with extreme paralysis, like individuals with locked-in syndrome.

1.2 Motivation

Playing music has been shown to provide several benefits for acquiring non-musical skills (Coffman (2002)). For instance, musicians have an improved ability to hear speech in noisy backgrounds (Parbery-Clark, 2009), reduced age-related auditory degradation (Parbery-Clark et al. (2011); Parbery-Clark et al. (2011)), increased verbal and auditory memory (Chan et al. (1998); Ho, Y. C., Cheung, M. C., Chan (2003)) and enhanced auditory attention (Strait et al. (2010)). Music instrument training is associated with neurostructural changes (Besson and Schön (2012); Wan and Schlaug (2010)) both in children (Hyde et al. (2009)) and adults (Bangert and Altenmüller (2003)). Motor brain regions are enlarged in musicians, when compared to non-musicians (Elbert et al. (1995)). Gray matter volumes tend to be larger in musicians than in non-musicians for motor, auditory and visio-spatial brain regions (Gaser and Schlaug (2003)). Furthermore, gray mater density is greater in Broca's (language) area for trained musicians (Sluming et al. (2002)). The corpus callosum, the fibers connecting the left and right hemispheres was found to be larger in musicians compared to non-musicians (Schlaug et al. (1995)). Musicians' resistance to age-related neural decline is greater for musicians when compared with non-musicians (Pagnoni and Cekic (2007)). Early instrumental musical training seems to train attentional networks in the brain, as well as social and interpersonal skills. Children exposed to musical training show improvements in nonverbal memory, IQ, numeracy and spatial cognition (Neville et al. (2008).

However, due to lack of fine motor skills, people with motor disabilities are often incapable of learning to play a musical instrument and thus, the benefits of music learning and performance are inaccessible to them. In

this context, Adaptive Digital Musical Interfaces (ADMI) provide a possible alternative for allowing people with motor disabilities to enjoy music playing and its associated benefits.

1.2.1   Types of Physical Disabilities

> Disabilities is an umbrella term, covering impairments, activity limitations, and participation restrictions. An impairment is a problem in body function or structure; an activity limitation is a difficulty encountered by an individual in executing a task or action; while a participation restriction is a problem experienced by an individual in involvement in life situations. Thus, disability is a complex phenomenon, reflecting an interaction between features of a person's body and features of the society in which he or she lives.[1]

Physical disability (or motor impairment) is a limitation on a person's physical functioning, mobility, dexterity or stamina. Other kind of disabilities include cognitive, intellectual, mental, sensory and developmental. Usually combination of those are present in individuals. The causes of physical disabilities vary, depending on the average income of the society. For example in India, the main cause of physical disabilities is Poliomyelitis infectious disease, when in developed countries such diseases are prevented through vaccination.

In this thesis, we are interested in the kind of physical disabilities that might limit an individual from playing music. In this section we introduce the type of disabilities that mainly concerned us in our research.

Paraplegic people are able to perform almost any musical instrument, as hand and torso movement is preserved. On the other hand, tetraplegia is paralysis caused by injury or illness that results in the partial or total loss of use of all four limbs and torso. It is caused by damage to the brain or the spinal cord at a high level (C1-C7). Typical causes of this damage are trauma (such as a traffic collision, diving into shallow water, a fall, a sports injury), disease (such as transverse myelitis, multiple sclerosis, or polio), or congenital disorders (such as muscular dystrophy). Sensory functions (touch, temperature, proprioception, pain) are normally also affected, de-

---

[1]World Health Organization. http://www.who.int/topics/disabilities/en/

pending on the cause of the physical disability. Tetraparesis, on the other hand, means muscle weakness affecting all four limbs.

Cerebral palsy describes a group of permanent movement disorders that appear in early childhood. In many cases it causes tetraplegia/tetraparesis, or even pentaplegia/pentaparesis. Pentaplegia/pentaparesis means all four limbs are involved, with neck and head paralysis often accompanied by eating and breathing complications.

Probably the most extreme case of physical disability is the locked-in syndrome. Locked-in syndrome is a rare neurological disorder in which there is complete paralysis of all voluntary muscles apart from the ones that control the movements of the eyes. The cognitive functions of people with locked-in syndrome are normally unaffected, but they have no ability to produce movements (outside of eye movement) or to speak. Communication is possible only through eye movements or blinking. In locked-in syndrome there is an interruption of all the motor fibers running from grey matter in the brain via the spinal cord to the bodys muscles and also damage to the centers in the brainstem important for facial control and speaking. Several different conditions might lead to locked-in syndrome, such as a blood clot (thrombosis) or stroke. Additional conditions that can cause locked-in syndrome include infection in certain portions of the brain, tumors, loss of the protective insulation (myelin) that surrounds nerve cells (myelinolysis), inflammation of the nerves (polymyositis), and certain disorders such as amyotrophic lateral sclerosis (ALS)[2]. Total locked-in syndrome is a version of locked-in syndrome wherein the eyes are paralyzed as well.

1.2.2 Objectives

In this PhD thesis, we will study three different types of interfaces designed for people with physical disabilities. In chapter 2 we present DMIs designed for people with limited upper-limb sensorimotor functions, in chapter 3 we present a gaze-controlled DMI, and in chapter 4 we present our research in brain-controlled music.

The objectives of the chapter 2 were to implement prototypes for people with different kind of physical disabilities and study how this influences their quality of life. The objectives of chapter 3 were (i) to build a gaze-controlled DMI, that would allow expressive music performances, (ii) to research on ways to optimize temporal accuracy in gaze-controlled musical

---

[2]http://rarediseases.org/rare-diseases/locked-in-syndrome/

interfaces and (iii) to implement and evaluate new gaze-selection algorithms that increase the spatial precision of gaze-controlled interfaces. Finally, the objectives of chapter 4 were to research on possible ways of (i) implementing Brain-Controlled interfaces that allow expression of simple musical concepts (like composing melodies or controlling the chord progression of a composition) (ii) research on emotion estimation methods using the EEG signal, and ways of using them in to enhance musical performances.

1.3 Contributions

This thesis investigates the implementation of DMIs for three main groups of people with physical disabilities: (i) limb-controlled DMIs, for people that preserve some degree of upper limb movement, (ii) gaze-controlled DMIs for targeted for people with locked in syndrome and (iii) brain-controlled DMIs targeted for people with total locked-in syndrome. The contributions of this thesis can therefore be summarized as follows:

Limb-Controlled

- Showing that using low cost materials, digital musical instruments designed for people with upper-limb paresis can be constructed and serve them for music expression and composition.

- Providing publicly available instructions and source code of how to implement the constructed prototypes.

Gaze-Controlled

- Implementation and evaluation of the EyeHarp, an open-source gaze-controlled music interface that allows expressive live performances.

- Implementation and evaluation of a new fixation-detection algorithm that provides improved temporal accuracy when compared to conventional fixation-detection algorithms.

- Implementation and evaluation of different gaze-selection techniques for improving the spatial accuracy in selecting small targets.

Brain-Controlled

- Introducing and evaluating a new method for emotion estimation using a low cost EEG device.

- Implementation of a hybrid Gaze-Brain controlled interface with emotion estimation.

- Showing that the performance of auditory Event-Related Potential based Brain Computer Interfaces is improved by introducing timbre, pitch and spatialization cues in the stimuli design.

- Evaluation of low-cost EEG devices in auditory Event-Related Potential based Brain Computer Interfaces.

- Public available database of recorded EEG signals in auditory oddball tasks using combination timbre, pitch and spatialization cues in the stimuli design.

- Implementation of 3 different auditory Event-Related Potential based Music Brain Computer interfaces aiming to allow basic music performance for people with total-locked in syndrome.

1.4   Thesis Outline

In chapter 2, we present three different studies. In each one of them we implement and evaluate an ADMI for a person with physical disability. In all three studies, the users maintain some degree of voluntary upper limb movement (upper limb paresis).

In chapter 3, we present and evaluate the EyeHarp, a gaze-controlled DMI. Additionally, a new fixation detection algorithm for improved temporal accuracy is presented and evaluated.

In chapter 4, we explore different ways of controlling music using brain activity. Initially, we report on an approach in which the brain activity measured in the motor cortex is used to control the contour of a melody. Then, we report on a proposed method for estimating emotional state of the individuals based on the measured brainwaves. We then describe a musical interface in which emotion estimation is combined with the previously presented EyeHarp interface. Then, we present a study in which low-cost Emotiv EEG device is evaluated in an auditory oddball paradigm, and another study in

which we explore how the timbre, pitch and spatialization auditory cues affect the performance of auditory P300 Brain-Computer Interfaces (BCIs). Finally, we present and evaluate three different Event-Related Potential (ERP) based Brain-Computer Musical Interfaces.

Finally, chapter 5 summarizes the contributions of our research.

Digital Musical Instruments for People with Limited Upper Limb Functioning

2.1 Background

In this chapter we will describe three different constructed prototypes made for three different cases of people with motor disabilities. In all three cases, the mental abilities of the persons were not affected. We are mostly interested in the kind of disabilities that prevent people from having access to traditional musical instruments. We will focus on cases of people with limited upper limb control. The main causes of such disabilities are cerebral palsy disorder and spinal cord injury.

In the past various musical interfaces for people with limited upper limb movement have been proposed and are in use.

Kirk et al. (1994) presented MIDICreator. The MIDICreator is a physical device with 14 sockets on the front panel, into which you can plug various different switches and sensors for detecting movements. It then sends midi messages to an external sound module.

Skoog[1], is a pressure and deformation sensitive cube. It can be played by pressing, squeezing, rubbing, stroking. tilting and shaking it. The mapping between the input and the output of the interface can be adjusted using the

---

[1]http://skoogmusic.com/, last accessed on 7th of July, 2016

accompanied software. Although it seems to be a user-friendly interface, it seems like it provides only basic interaction. It might me an appropriate interface for specific users.

Bhat (2010) developed TouchTone, a tangible interface with 10 keys arranged in two rows. The keys are big enough to be pressed by children with cerebral palsy. The interface was qualitatively evaluated on children with cerebral palsy. According to the author, the instrument was found to offer a very high playability across variations in physical ability.

Swingler (1998) is the creator of SoundBeam. Soundbeam is a commercial product [2]. Soundbeam utilizes a sonar proximity sensor and detects the distance from the sensor of a chosen part of the body of the user. The distance and velocity of the chosen part of the body (e.g. hand, head) is translated into midi messages controlling an external sound module.

Similar computer vision approaches have been proposed, such as the Movement to Music System (MTM) introduced by Lamont et al. (2000). The system uses a web camera to detect movement in certain regions of the captured frame. Every region is assigned to a specific sound or note. These regions can be set by the user. Tam et al. (2007) reports that the system has been found to have positive impact in the body functions of children with limited movement, while it allows them to have independent play activities. One of the advantages of this system is its low cost. Another even more flexible similar system was recently proposed by Oliveros et al. (2011).

In sections 2.2 and 2.3 we describe the implementation and evaluation of two prototypes made for two persons with spinal cord injury. In section 2.4 we describe the implementation and evaluation of a prototype made for a person with cerebral palsy.

2.2 A percussive midi controller for a quadriplegic person

2.2.1 Introduction

In this section we will describe the construction of a midi-controller for a quadriplegic person. All mentioned interfaces in the previous section are designed mostly for therapeutic purposes. They do not allow the expression of complex melodies. Moreover they do not allow expressive performances.

---

[2]http://www.soundbeam.co.uk/

In this section we describe the implementation of a prototype constructed for an amateur musician that after a spinal chord was left quadriplegic. The purpose of the prototype is to provide him with a midi controller that would allow him to play music in a similar way traditional musical instruments are being played.

The subject did not have any objection in revealing his identity. His name and surname is Kostas Manikis, born in 1982. He was born and currently lives in Thessaloniki, Greece. As a child, he had 5 years of musical training on the keyboard but he abandoned it. At the age of 18, as a student at the Department of Music Technology and Acoustics Engineering in Rethymno, Crete, he started playing the acoustic guitar as self-educated. At the age of 25 he starts playing the guitar in local bands of rock, jazz and Greek traditional music, making a small income out of his performances. At the age of 28, just before finishing his Bachelor studies, and while being an active musician, he had a car accident, in which he was injured in the spinal cord, at the C6/C7 region. Since then he is quadriplegic. Particularly from the chest and below all motor-sensory functions are impaired. The movement of the arms is limited. All movements are preserved till the elbows, apart from the triceps in both hands. From the elbow and under no motor movement is preserved in both hands. The sensory functions of both arms are also limited. In order to keep his balance, he normally has to hold from the wheel chair. Kostas can control the computer using a trackball and a normal keyboard, after adapting to his hands a special glove with a small stick at the bottom.

After the injury he bought and tried to play the santouri, a Greek traditional stringed instrument in the hammer dulcimer family (see figure 2.1). The mallets were tied on his hands. From the beginning though it was clear that this instrument was not appropriate for his situation. The santouri consists of 72 chords. Tuning the instrument is a time-consuming task that requires high precision even for musicians without disabilities. Kostas was dependent on friends musicians visiting him in order to tune the instrument for him. Another limitation of the instrument was the fact that it required high spatial precision, due to the small distance between the cords. Finally reaching the distant chords was impossible, as this resulted in loss of equilibrium of the body, due to paralysis of all muscles below the level of the chest.

Kostas contacted me, asking me if I can help him implement a music controller that he would use for composing and performing music. The process
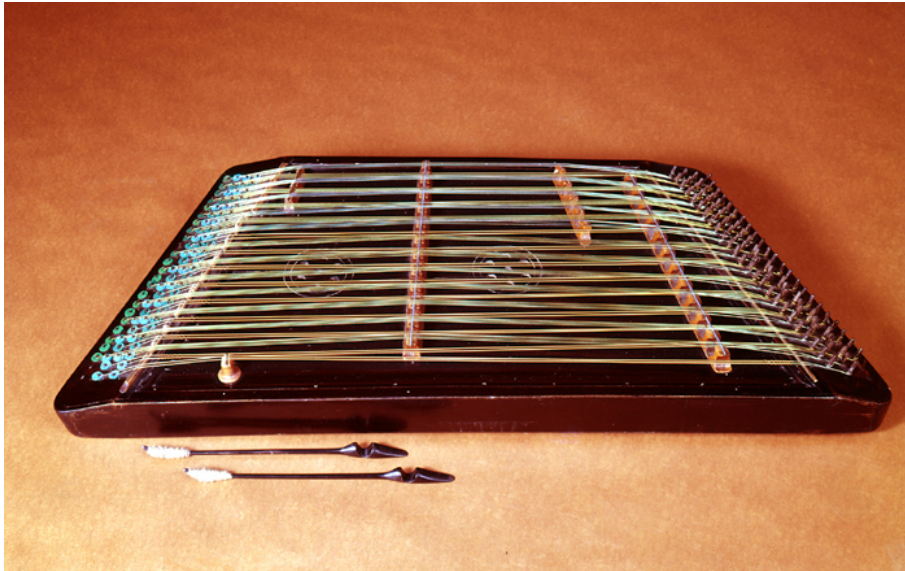
Figure 2.1: The Santouri.

of designing a musical interface for Kostas was a cooperative task. Kostas is a person with education in both music and music technology. Our purpose was to design and construct an interface that would overcome the limitations introduced by his disability. Nevertheless, it should encourage him make use of his motor abilities. Additionally it should work as a midi controller, something that would allow him to use a big variety of sound modules on his computer, record his musical ideas and compose music.

2.2.2 Materials and Methods

After trying different mock-ups, we agreed on implementing a percussive interface that would be played using mallets. The available buttons should be placed in a reachable distance. Figures 2.2 and 2.3 show the first constructed working prototype.

It consisted of 17 piezoelectric sensors, connected to the analog inputs of an Arduino uno micro-controller. As the available analog inputs of an Arduino uno are only 6, a Mux Shield II from Mayhew Labs [3] input expander was used, providing 48 available analog inputs. After uploading the software to arduino uno, the HIDUINO firmware (Diakopoulos and Kapur (2011))

---

[3]Available Online at http://mayhewlabs.com/products/mux-shield-2

Figure 2.2: First prototype.

is utilized, in order to transform arduino into a true USB-MIDI device for plug-and-play compatibility.

The piezoelectric sensors, also known as knock-sensors, produces a voltage in response to some type of physical stress, such as a knock. Each sensor was connected in parallel with an $100k\Omega$ resistor and an analog input of the shield as shown in figure 2.7. In order to increase the surface of each knock-sensor, it was glued to a round plate made of iron of diameter 6 cm. The iron plates were cut using a sheet metal scissor. On top of the iron plates, a layer of 0.6 cm PVC foam was glued. Each of these sensors were then placed on 5 cm thick PVC frame of dimensions 50 x 25 cm. This first prototype was constructed on January, 2014. Three months later, a more robust construction was made. The PVC frame was replaced by a wooden one. Another 1-cm thick PVC foam disc was placed under each knock sensor. An MPX5010DP blow sensor (figure 2.6) was also added, mapped to the expression midi control-message. Figures 2.4 and 2.5 show the final prototype. The wooden body of the prototype was with the help of Dimitris Fotopoulos, who professionally constructs musical instruments, like guitar and bouzouki [4].

---

[4]https://fotopoulos.wordpress.com/
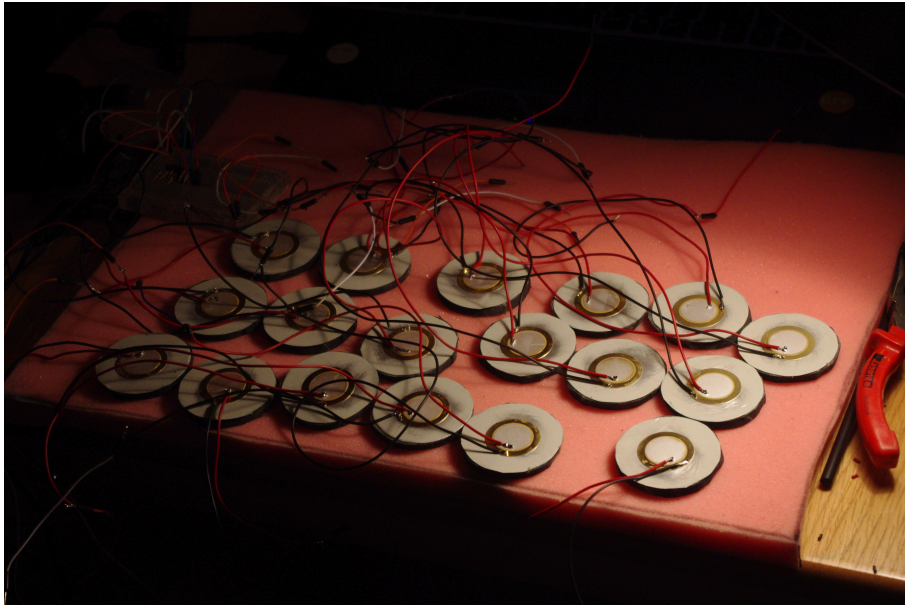
Figure 2.3: Knock Sensor of the first prototype, before pasting them on the base made of foam.

A simple threshold was used to detect whether a sensor was hit. This threshold was set individually for each sensor.

The arrangement of the notes is chromatic, just like they are arranged in a chromatic accordion, known as bayan. The sensors are arranged in 3x6 arrangement. The notes of the leftmost column are 3-c (c in the 3rd octave), 3-c#, and 3-d. The notes on the second leftmost column are 3-d#, 3-e and 3-f. Moving up this way the notes at the rightmost column are 4-d#, 4-e and 4-f. The range of the controller is 3-c to 4-f.

An additional knock-sensor button is placed at the top-right corner of the interface. This button is is used as a multiplexer switch. It is used for switching between the "settings" and the "melody" mode. In the melody mode, each button corresponds to a note, as described above. In the settings mode various parameters of the controller can be set. Table 2.2.2 shows all available settings.

The interface might be monophonic or polyphonic. In the case it is polyphonic, a preset duration value is set for each note. The default value is 0.5 seconds, and the user can change that value in steps of 100 ms. When in

Figure 2.4: Final prototype.

the polyphonic mode, through the blow sensor, attached to the mouth, the user can control the velocity of the note-on midi messages. In the monophonic mode, the expression control message is controlled through the blow sensor. Both in the polyphonic and monophonic mode, a way to release all notes is by blowing out. The expression or velocity values are controlled by blowing in. The functions of blowing in and out can be inverted by placing the mouthpiece of the blow sensor in each of the two available holes of the sensor (see figure 2.6), as the one hole measures positive pressure and the other one measures negative pressure (vacuum). As mouthpiece we refer to a small straw. In the monophonic mode, a note is released either when another note is played or when the user blows out (when blowing-in controls the expression control message). If the blow sensor stays receives no significant fluctuations for a period more than 5 seconds, the expression and velocity take a preset value. This allows the user perform without using the blow sensor. In that case though he is not able to control the dynamics of the performance. The blow sensor is hanged around the neck of the user, with the straw being close to his mouth.
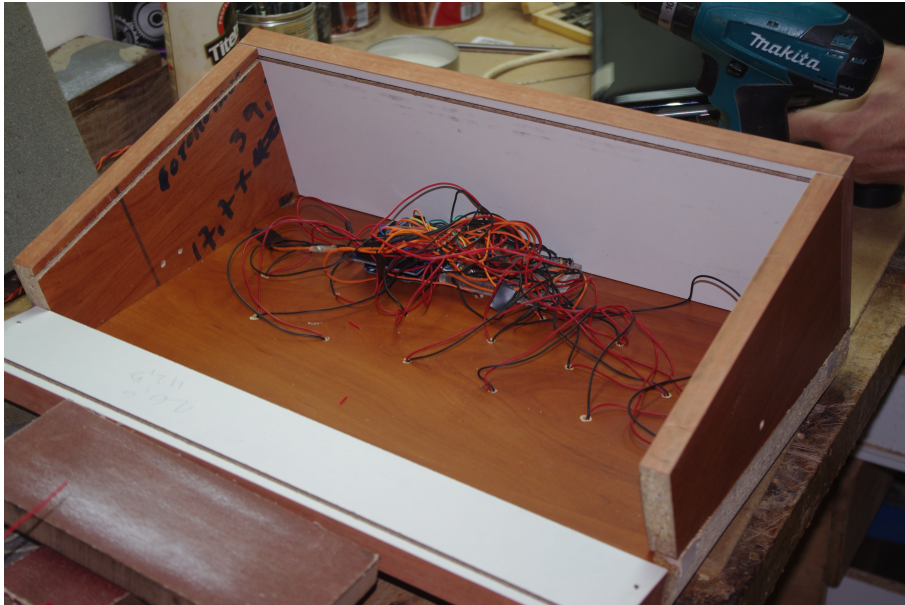
Figure 2.5: The inner part of the prototype when opened.

Corresponding note	Setting
3-c	Transpose one octave down
3-c#	Transpose one octave up
3-d	Switch Monophonic-Polyphonic Mode
3-d#	Decrease Note Duration (-100 ms) in Polyphonic Mode
3-e	Increase Note Duration (+100 ms) in Polyphonic Mode
3-f	Reset all Settings
3-f#	Transpose one semitone down
3-g	Transpose one semitone up
3-g#	Decrease midi channel number
3-a	Increase midi channel number

Table 2.1: Available Settings.

2.2.3 Evaluation

In order to quantitative evaluate the constructed prototype, Kostas was
asked to make a composition and study it. In a period of 4 weeks, at the
end of each week, we made a recording of the performed melody. Only
during the first week, also studied and performed the same melody using
another midi controller that he had been using for about a year along with

Figure 2.6: NXP MPX5010DP Blow sensor.
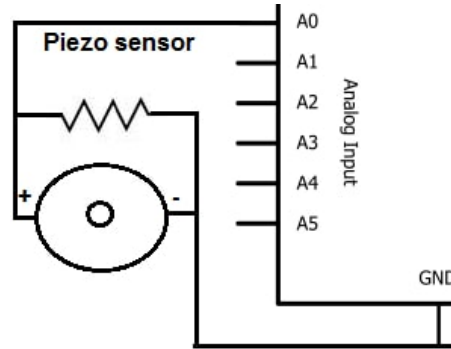
Figure 2.7: Schematic for connecting each piezoelectric sensor to the analog input expanded shield attached to the arduino uno micro-controller.

the constructed prototype, the korg nanokey 2 midi keyboard (see figure 2.8).

The temporal accuracy and the number of mistakes (omitted notes, wrong notes) were measured. The score of the composed and performed melody is shown in figure 2.9.

2.2.4 Results

Figure 2.10 shows the Box and Whisker chart of temporal asynchrony of all 5 recording. A Box and Whisker chart indicates the median value of a dataset and ts variance. The line (Q3) at the middle of each box corresponds to the median value of a set. The lower end of the box (Q2) is the median of the lower half of the dataset, and the upper end of the box (Q3) is the median of the upper half. The lower 'T' symbol corresponds to the lower value of the dataset, and the upper to the higher. Outside these areas appear the outliers as small dots. Outliers are the values more than 1.5 times the distance between Q1 and Q3.

Figure 2.8: The Korg nanokey 2 is a small midi controller with total length 32.5 cm.

Figure 2.9: The music score of the composed melody that served for the quantitative evaluation of the prototype.

	Korg Nanokey	Week 1	Week 2	Week 3	Week 4
Omitted Notes	0	3	5	6	1
Wrong Notes	3	0	0	0	0

Table 2.2: Number of wrong and omitted notes for each recording.

Table 2.2.4 shows the number of wrong and omitted notes of each performance. As omitted we refer to a note that was not played, and as wrong to a note that did not exist in the score.

2.2.5   Discussion

Looking at figure 2.10 there is no evidence of improvement in time regarding the temporal accuracy of the performance. In general the melody was per-
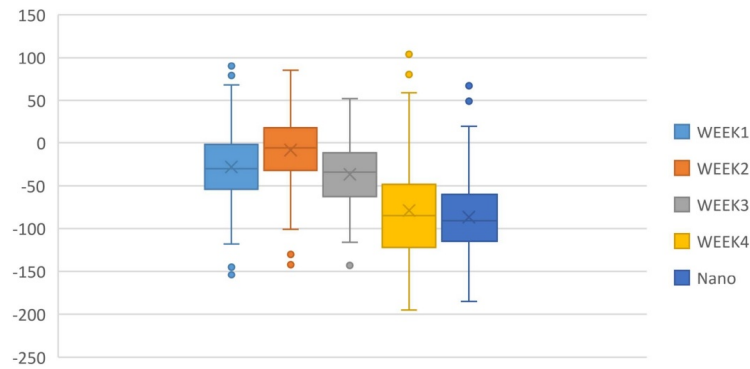
Figure 2.10: Box and Whisker chart of temporal asynchrony of each of the 4 recordings.

formed well starting from the 1st recording. This indicates that probably a more demanding task should have been used for studying his improvement.

As seen in table 2.2.4, when using the constructed prototype, some notes are not being played, while when using the Korg nanokey midi controller, additional, wrong notes are played. These two facts reflect the weaknesses of each controller. The Korg nanokey has small buttons, so it is prone to pushing adjacent buttons. Nevertheless due to its small size it is more transportable. On the other hand, when performing with the constructed prototype no wrong notes are played. Nevertheless, some notes are omitted. This indicates that the sensibility of some buttons should be lowered (i.e. the threshold value in Volts that should trigger an event).

The final prototype was implemented and given to Kostas at the mid of April, 2014. Since then Kostas has been playing music in a daily basis, alone and with other musicians using the constructed prototype. He also got involved in composing music. He reports that when jamming with other musicians he prefers the constructed prototype over any other controller that he have tried so far. The reason is that we were able to adapt the interface to his needs. As Kostas is incapable of moving his fingers, the only way to play legato would be through a percussive instrument, in which the release time is predetermined and can be adapted before the performance. Or, alternatively by using a blow sensor to control the dynamics, the release and in general the expressive characteristics of the performance.

Advantages	Disadvantages
"It has big buttons that favor spatial accuracy"	"It is heavy. Although this was my choice when designing it. I now think that because of this it lacks in mobility."
"It has different arrangement of the notes when compared to a piano. This helps in composing and playing melodies that most probably I wouldn't in a conventional controller"	"I cannot update its firmware without the help of a non-disabled expert. For that reason I have not used the breath controlled that much."
"It offers a lot of settings that can be easily accessed, such preset note duration in the polyphonic mode"	"It is easy to forget which button corresponds to each setting"

Table 2.3: Advantages and disadvantages of the constructed prototype as reported by the user.

On his soundcloud channel [5] there are currently 7 musical compositions, 3 of which were recorded using exclusively the constructed prototype. Alternatively, he inputs the notes one by one, using a trackball, or korg nanokey 2 midi keyboard.

He was asked to mention the advantages and disadvantages of the constructed prototype with conventional midi keyboards he mentioned the advantages and disadvantages shown in table 2.2.5.

The prototype can be improved in many ways. Its weight should be reduced. This would make Kostas able to move it from his desk to his legs in order to properly play. Although the prototype was constructed in order to play with both hands, after placing it on his feet, because of its weight, it was difficult for him to place it in a proper position close to his body. Consequently, most of the time he plays with one hand, as he used the other one to hold from the chair.

Every time a firmware update is required, a shortcut should be made between two pins on the Arduino board, in order to re-install the arduino firmware and then update the new software. Then another shortcut is made between the same pins in order to re-install the HIDUINO firmware, which enables the Arduino to be recognized as a midi controller. Consequently,

---

[5]https://soundcloud.com/between2notes

in order to make any software update, we had to travel from Barcelona
to Thessaloniki. Unfortunately, after adding the blow sensor we were able
to visit Kostas just 1 more time. Although tests were made to determine
the optimum sensitivity of the blow sensor, in the long term it was proven
that the sensitivity was low. Too much effort was required to control the
dynamics of the performance. On our next visit to Thessaloniki, an update
will be performed, which will allow adjusting the sensitivity of the sensor
from the settings layer.

In summary, we constructed along with Kostas a percussive midi controller
designed for him. Although the prototype has some disadvantages, like its
weight, it allows Kostas to play music with other musicians, and compose
music. He mostly uses the controller for improvising or playing known
traditional greek songs. Since the prototype was constructed, he plays music
in a daily basis and it has motivated him to get involved more in music
activities. Recently he composed music for the theater and he sees his
future as a semi-professional composer.

2.3   A Guitar Controller for a Person with limited Sensorimotor Functions

2.3.1   Introduction

In this section we will report on the implementation of a guitar controller
for a person with Brown-Squard syndrome. The work presented in this
section was conducted in coordination with Joan Sandoval and forms part
of his bachelor final thesis project (Sandoval Codina (2015)). Brown-Squard
syndrome is caused by damage to one half of the spinal cord, resulting in
paralysis and loss of proprioception on the same side as the injury or lesion,
and loss of pain and temperature sensation on the opposite side as the lesion.
Proprioception is the sense of the relative position of neighboring parts of
the body and strength of effort being employed in movement.

Our subject is a 38 year old women, named Laura. Laura suffers from
tetraparesis. Symptoms appeared after the extraction of a tumor in the
spinal cord 5 years before we met her. At the time we were implementing the
music controller for her, she was a music therapy student. The movement
of right hand is limited with the fingers being paralyzed. She has lost the
sense of touch and temperature in the left part of his body in all regions
under the neck.

Before the surgery, she could play the piano and the guitar without diffi-
culties. She now can make only simple melodies in the piano using the left
hand. With the guitar she can play chords and strum, accompanying her
voice, but she is slow placing the fingers on the right positions, as she lacks
the sense of touch. In both cases, she has to look at her hand when play-
ing, as her proprioception is lost. Laura wants to become a music therapist
working with children. The fact that she has to look at her left hand all the
time, limits her communication as a therapist with the children.

2.3.2   Materials and Methods

Preliminary Prototypes

From the beginning Laura expressed her to desire to be able to play the
guitar without having to look at her left hand while playing. She agreed
that we should make a guitar controller that would allow her to play different
chords, either by strumming or by pressing the notes one by one in order to
make arpeggios. Different initial prototypes were tried before implementing
the final prototypes.

The first prototype was implemented using the leap motion[6] device, capable
of capturing hand and finger movements in three dimensions. In this first
prototype we focused on how melodies and arpeggios could be played.

The software part was implemented using processing programming lan-
guage. As Laura is not capable of moving the fingers on the right hand,
arpeggios could only be played with the left hand. As seen in figure 2.11,
in the first prototype the position of the right hand determined the selected
chord, while with left hand, the different strings can be played. In figure
2.11, in the left part of the screen the white points correspond to the de-
tected fingers. The Leap motion sensor is placed on desk. The vertical
movement of the fingers corresponds to the distance from the sensor. If
each of the fingers crossed the upper blue line, the note corresponding to
this finger was played, depending on the selected chord. If the line placed
lower was crossed, the same note one octave down was played.

When Laura tried the prototype, although she was able to accurately play
arpeggios, she complained that having both hands lifted in form of her for
a prolonged time would be tiring. Moreover she was dependent on looking
at the screen in order to know where to place her hands.

---

[6]https://www.leapmotion.com/, last accessed on the 10th of July, 2016.
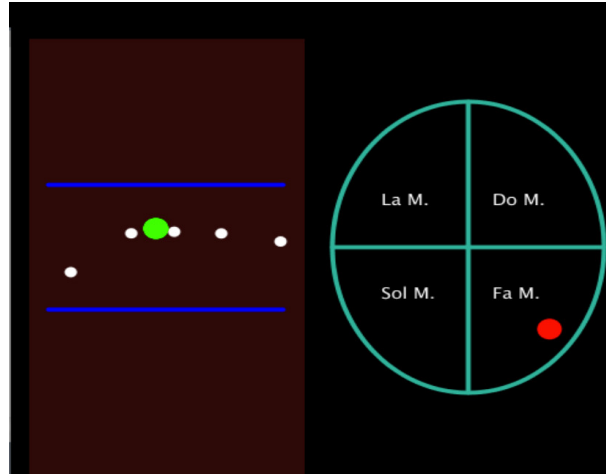
Figure 2.11: Visual feedback of the first prototype based on Leap motion sensor.

In the next implemented prototype, the arduino microcontroller was utilized
along with an accelerometer and a string-pot sensor. A string potentiometer
is a transducer used to detect and measure linear position and velocity using
a flexible cable and spring-loaded spool. For the needs of this project two
low-cost string pots were constructed. The body of the string-pots were
printed in a 3d-printer. The designed parts fed to the 3d printer were
provided by Chap Robotics FRC team 2468. [7]

Figure 2.12 shows the second implemented prototype. The user holds the
device on the left hand with the fingers placed on each of the 4 buttons.
For selecting the chords, the distance measured by the stringpot was used,
along with the inclination of the construction.

When the subject tried the prototype, it was clear that the inclination
could not be used as a parameter. The loss of proprioception on the left
hand resulted in inaccurate control of the inclination. Nevertheless, the
use of stringpots seemed to be a good option for tracking the movement of
the right hand, as it worked in a robust way and it can provide a visual
feedback to the user. If the string is painted with different colors, the user
has feedback on the length of the string and as a result the selected input
action.

---

[7]Stringpots STL and assembly instructions available online at
http://www.andymark.com/product-p/am-2618.htm, last accessed on the 10th of
July, 2016.
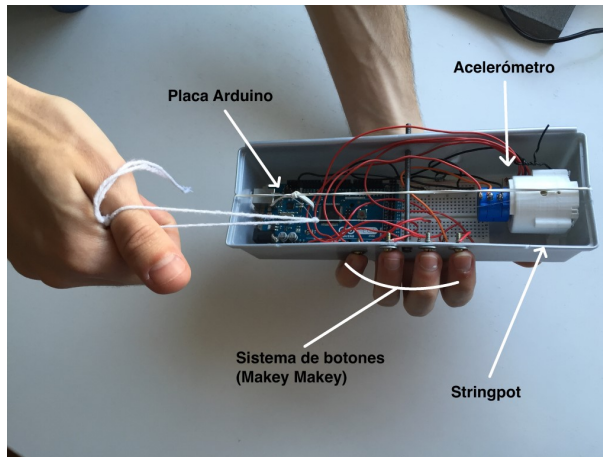
Figure 2.12: The second prototype.

The final prototype

The final prototype was built by modifying the guitar hero III controller made for Nintendo Wii gaming console. The main board of the controller was replaced by the arduino microcontroller. As in 2.2, the hiduino firmware was used, enabling the arduino microcontroller to communicate through the midi protocol with a sound module running on a computer. Two stringpots were also mounted inside the enclosure of Guitar Hero. The buttons of guitar hero were connected to the digital inputs of arduino. Figure 2.13 shows the back of the modified guitar hero, while 2.14 shows the front of the controller. Figure 2.15 shows the subject playing the final prototype.

Using the two stringpots, it was possible to compute the position of the ring that joins them. If a and b are the distances computed by the two stringpots, then by applying the Pythagorean theorem, the calculated the coordinates (x,y) of the position of the ring are:

$$x = \frac{a^2 + c^2 - b^2}{2 \cdot c}, y = \sqrt{a^2 - x^2}$$

,

Tracking the right hand's movement in 2 dimensions was used for strumming virtual strings. The virtual strings were equally distributed close the the guitar's body, the same way they are placed in a normal guitar. Every
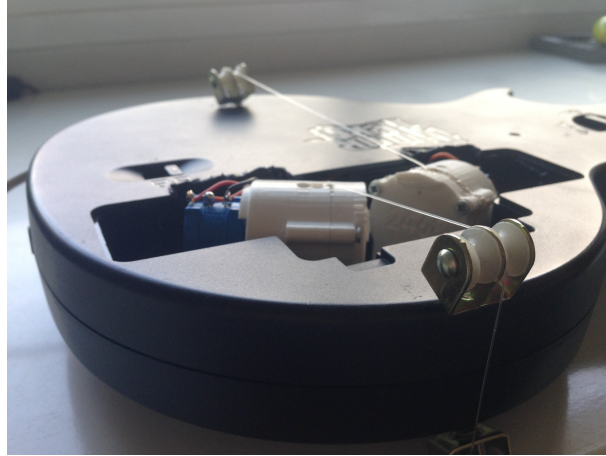
Figure 2.13: The back of Guitar hero with modifications, including two stringpots
and pulleys.
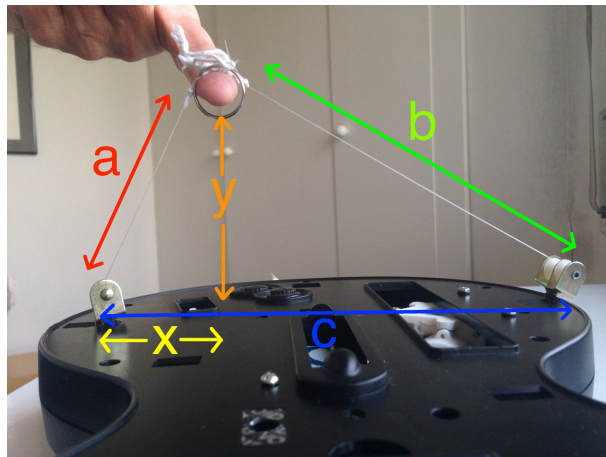
Figure 2.14: The front part of Guitar hero with modifications. The position
in two dimension of the hand of the user can be computed by considering the
measured distances of the two stringpots.

time the tracked position crosses a string, the corresponding noteON midi
message is send.

The chords are selected by pressing combinations of the 5 buttons placed
on the neck of the guitar. Figure 2.16 shows the combination of buttons
used to play each of the 7 available chords.

Figure 2.15: Laura playing the final prototype.
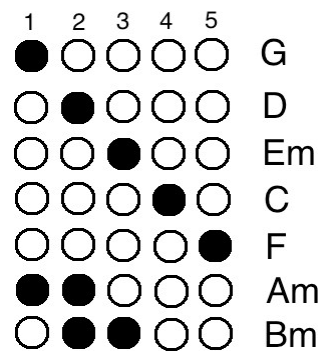
Figure 2.16: Combination of buttons for playing different chords.

2.3.3 Evaluation

Quantitative

Once the prototype was ready, it was given to the subject and in a period of a month 7 recording were made. The given task is shown in figure 2.17. Before making every recording she practiced for 10 minutes the given task.

Figure 2.17: The chord progression along with the strumming pattern given to
the subject for evaluating the prototype. The arrows show the corresponding hand
movement when strumming.

The recordings were made without using a metronome. When analyzing the
results, in each recording the mean tempo was computed. We report on the
temporal accuracy of the performed strums and on the percentage of these
strums played correctly. The temporal accuracy, was measured by counting
the percentage of strums played in-tempo. In-tempo was considered every
strum with temporal accuracy less or equal to one sixteenth note. The
accuracy in played the right chords was measured by subtracting from the
expected number of chords the number of mistakes. A mistake is counted
in two cases: (i) if during a strum less than two strings were played, or (ii)
if the wrong chord was selected when strumming.

User experience

After a month of using the interface, the subject commented on her expe-
rience with it, the problems that appeared at the moment of using it and
proposed ways to improve it.

2.3.4 Results

Figure 2.18 shows the average temporal accuracy and percentage of strums
played with the correct selected chord across the sessions.

Regarding the user experience evaluation, the most important problem of
the interface was the high resistance of the stringpots. Additionally they
produced a sound when playing the instrument. This was a problem when
playing at low amplitude. On the contrary she found convenient the use
of the buttons on the left hand used for select chords. The subject also
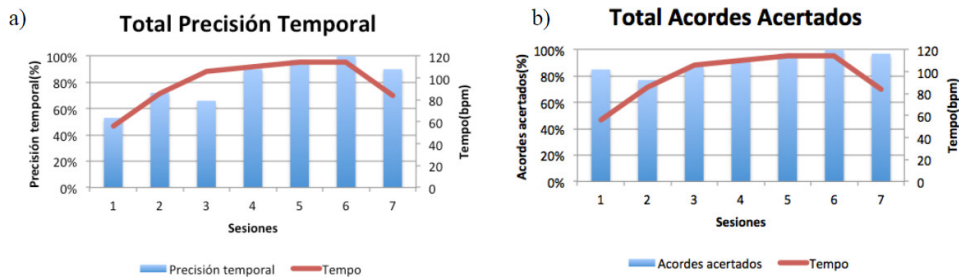
Figure 2.18: Average temporal accuracy and percentage of strums played with the correct selected chord across the sessions.

complained about the sound quality of the system. Nevertheless this aspect was not considered in this study. Using a different guitar synthesizer would improve the sound quality of the prototype.

2.3.5 Discussion

As seen in figure 2.17a, after the third session there is a clear improvement in the temporal accuracy of the strums. In the sixth session a perfect performance was recorded. While in the first session the temporal accuracy score 50%, in the last session it is 90%. A slight improvement in the performance is also observed in the chord accuracy. In the first session it was 82% and in the last session 97%.

The usage of stringpots for detecting the hand movement was proven to provoke serious negative impact in the user experience because of the high resistance and the produce noise when playing. Probably another approach should be considered. In the following section, in figure 2.19 is shown an implemented prototype for testing various input devices. One of them is a series of touch-sensitive strings. This would probably be a better approach to the problem of strumming. Nevertheless, as the subject suffers from lost the proprioception and sense of touch on the left hand, the appropriateness of such an approach should be tested. The subject is prone to injuries, as she lacks the feel of pain on the left hand.

2.4   A Guitar Controller for a Person with Cerebral Palsy

2.4.1   Introduction

The work presented on this section was mainly performed by Schmidt (2014) under my co-supervision with Rafael Ramirez. In this section a review of the mentioned master thesis will be maid. All software and hardware of this project was implemented by Marcel Schmidt under my co-supervision. Figures and tables are copied from his master thesis document.

The most recent definition of Cerebral Palsy, based on modern brain imaging techniques, was given by Rosenbaum et al. (2007) :"Cerebral palsy (CP) describes a group of permanent disorders of the development of movement and posture, causing activity limitation, that are attributed to non progressive disturbances that occurred in the developing fetal or infant brain. The motor disorders of cerebral palsy are often accompanied by disturbances of sensation, perception, cognition, communication, and behavior, by epilepsy, and by secondary musculoskeletal problems".

Cerebral palsy is the most common disability in children, rating between 2 and 3 per 1000 live births.

In this case study, the subject is a 36 years old man from Germany. He suffers from dyskinetic cerebral palsy with choreoathetoid which is characterized by irregular, twisting and curving movements. According to the Gross Motor Function Classification System (Palisano et al. (1997)), the degree of his disability falls between categories 4-5 out of 5. He is not able to walk. His disease prevents him from eat or drink without assistance. He speech is also affected. He lives in a home for people with motor disabilities.

His mental abilities are not affected, and he finished secondary school. He works in a sheltered workshop writing articles. He controls the computer using a trackball and a special keyboard. He was strongly motivated to learn playing a musical instrument with a preference over the guitar. In the past he had tried to play the keyboard. Nevertheless, he never had any proper music training.

From initial interviews, the subject expressed the desire to play the guitar. In the following section we describe the implementation of a guitar controller adapted to his needs.

2.4.2 Materials and Methods

Determining the final Prototype

The initial design of the system was performed considering the movement abilities of the user and the nature of the guitar as an instrument. The first priority was to design a system that would allow changing the chords of the guitar and strumming. The second was to provide the potential of playing melodies. For accomplishing the first goal, the input system should have at least 2 analog degrees of freedom: one for selecting the chord and a second one for strumming.

The user should be able to select a set of N chords in a specific order. Then using one of the analog inputs, he should be able to select the desired chord. This would be achieved by dividing the range of the analog input into N equally distributed areas, each of them assigned to a chord. If for example the selected chords were F, C, G, Dm and Am, then the range from 0% to 20% would correspond to the chord F, the range 20% to 40% to the chord C and so on. The same idea was be applied for strumming the strings. The number of virtual string was six, as in a normal guitar. The first string was played when the analog input crossed passed the value corresponding to the 10% of its range. The last one when the analog input crossed the 90% of its range. The remaining string were equally distributed between the first and the last string.

An initial version of the software was implemented in which the strumming and chord selection functions were implemented. The software processes the user input, converts it to MIDI data and sends it to the connected software synthesizer. The used synthesizer was the VirtualMIDISynth from CoolSoft on a Windows computer or QSynth on a Linux computer. The software was implemented using processing programming language[8]. A first prototype for testing these functions was implemented, including a variety of sensors. All sensors were connected to arduino uno micro-controller. Figure 2.19 shows the first prototype implemented for testing different kind of sensors for controlling either of the two parameters of the system. Additionally the EyeTribe [9] tracker was tried as a gaze-input device, and the camera mouse (Betke et al. (2002)) as a head tracking system. All inputs were tried for each of the strumming and chord selection functions for about two minutes. An additional person was observing the procedure. At the end, the subject

---

[8]https://processing.org/, last accessed on 8th of July, 2016
[9]http://theeyetribe.com/, last accessed on 8th of July, 2016.

Figure 2.19: The first prototype used for testing a variety of input methods

and the observer an were asked to rate each input device in a scale from 1 to 5. Table 2.4 summarizes the ratings of both the observer and the user. As guided hand movement we refer to a movement in which the subject was physically restricted to move an object in only one dimension.

It was impossible to achieve a reliable tracking of the gaze of the subject. As seen in table 2.4, both the subject and the observer agreed that the finger movement using the joystick was the most appropriate input for strumming, while the left to right arm movement was judged as the most appropriate input for fretting.

We then decided to compare the joystick over a scroll wheel. The user preferred the mouse wheel over the thump joystick as it "gave more control and felt more natural as there were small bumps, like real strings". A custom scroll wheel was constructed, in which a force resistive sensor was added, making possible the measurement of the force applied to the scroll wheel, that could be used for controlling the velocity of the notes.

The final Prototype

Figure 2.20 shows the final prototype. It consists of a "sledge" that can be freely movement in one dimension. On the "sledge" there is a scroll wheel and a button. An ultra sonic proximity sensor placed at the left edge of

Movement (Sensor)	Strumming		Fretting	
	User Rating	Observer Rating	User Rating	Observer Rating
Finger movement (buttons)	3	2	3	3
Finger movement (thumb joystick)	**4**	**5**	2	1
Finger movement (touch-sensitive strings)	3	2	N/A	
Hand movement (slide potentiometer)	2	2	2	3
Arm movement (proximity sensor)	1	1	1	1
Guided arm movement (proximity sensor)	2	3	**5**	**5**
Head movement (Camera Mouse)	N/A		2	3
Eye movement (EyeTribe)	N/A		N/A	

Table 2.4: User and Observer Rating for different input methods

Figure 2.20: The final prototype.

the device outputs the distance position of the sled from it. The device is connected through a usb port to a computer running the implemented software. The software, implemented in processing programming language, translates the received data to midi messages, driving a guitar synthesizer. It also provides the visual feedback on the computer's screen. Figure 2.22 shows the visual feedback provided when playing the instrument. A detailed overview of the software was presented by Schmidt (2014).

The system provides auto-strumming, auto fretting functions and supported strum functions. When auto-strumming is ON, the system will strum automatically following a preset pattern specified in a setting menu. At the same menu, the user enters a preset chord sequence. When auto-fretting is ON, the chords will be changed automatically, following the defined sequence. When supported strum function is turned on, every-time the user scrolls, a full strum is assured on the moved direction.
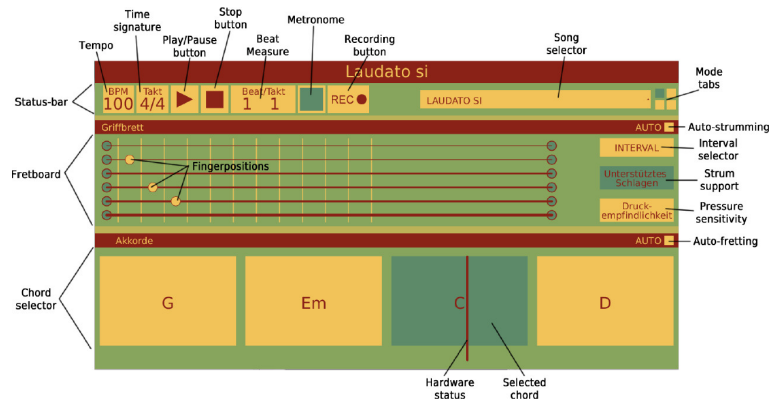
Figure 2.21: The graphical user interface in play mode.

2.4.3 Evaluation

Quantitative

After the prototype was implemented it was given to the subject so he could use it on a daily basis and practice playing it. Figure 2.22 shows the subject playing the final prototype.

A quantitative evaluation was conducted in order to measure the improvement with the interface in time, on tasks of increasing difficulty. In that case the subject was asked to record 3 different song of increasing complexity once a week in period of 9 weeks. Each song was recorded for approximately one minute in each mode: first with auto-strumming support, then with auto-fretting support and last without any support. Table 2.5 shows an overview of the selected songs.

For each recording the missing chords, additional chords, wrong chords and correct chords features were extracted. Additionally, the mean distance and standard deviation between the position of the correctly played chords and the the supposed positions of the chords were calculated. From each recording, 90 beats were extracted for analysis. We report in the amount of strums (chords) performed by the subject (CA), additional chords played (AC), missing chords (MC), wrong chords (RC) and correct chords (CC). The AC value is the number given when subtracting the CA from the beat amount (=90). The MC values indicates how many of the supposed chords were missing.

Figure 2.22: The subject using the guitar controller.

Song Name	Die Welt	Laudato Si	Hallelujah
Tempo in bpm	80	80	80
Chords	A, D	G, Em C, D	G, D7, Am, C
Complexity	Low	Intermediate	High
Strum pattern	Down-Down-Down-Down		
Time Signature	4/4		

Table 2.5: The used in the quantitative evaluation.

Usability Test and User Experience

The usability and user experience test aims to provide information about the subject's ability to achieve given goals and how he experiences the interface. For the usability test the subject had to accomplish a list of given tasks. While doing those tasks he was observed and asked to think aloud. After all tasks were accomplished the subject was asked to freely explore the interface and was interviewed afterwards. In the interview he was asked to tell the

(a) Auto Strumming

(b) Auto Chord
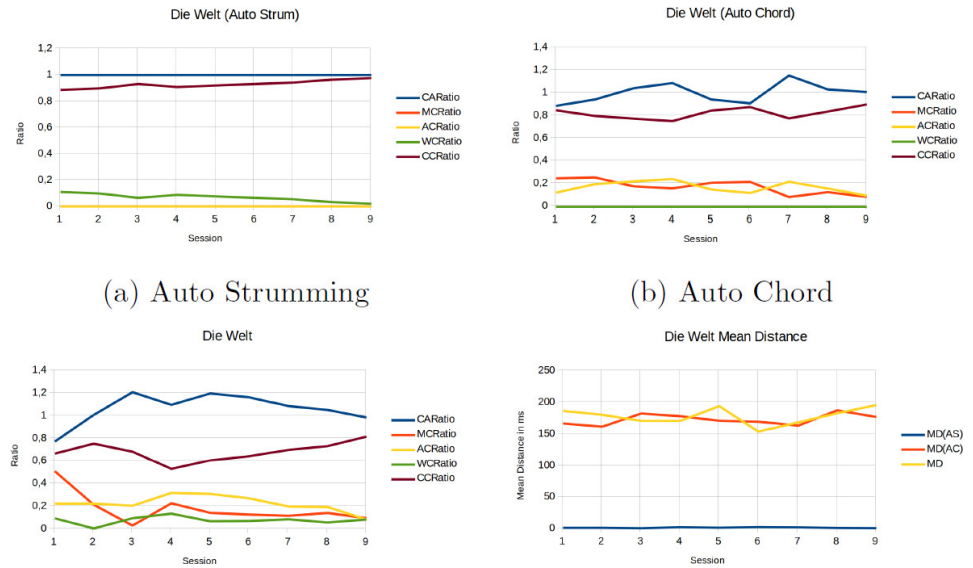
Figure 2.23: Results for the song "die Welt".

interviewee about his experience while using the interface as well as how the interface affected his life in general.

The usability test does not aim to test how well the user can play a song with the interface but rather if he can achieve goals such as setting up a song or changing the MIDI port. It consisted of 39 tasks, such as creating and naming a new song, setting the tempo and the time signature, adding chords, activating auto-strumming, playing only the first 3 strings of the guitar and closing the software.

For the user experience test an unstructured interview was chosen to let the subject freely express his feelings without giving him the impression of being interviewed.

2.4.4   Results

Quantitative

Figures 2.23, 2.24 and 2.25 show the results for each song over all nine sessions.

(a) Auto Strumming

(b) Auto Chord
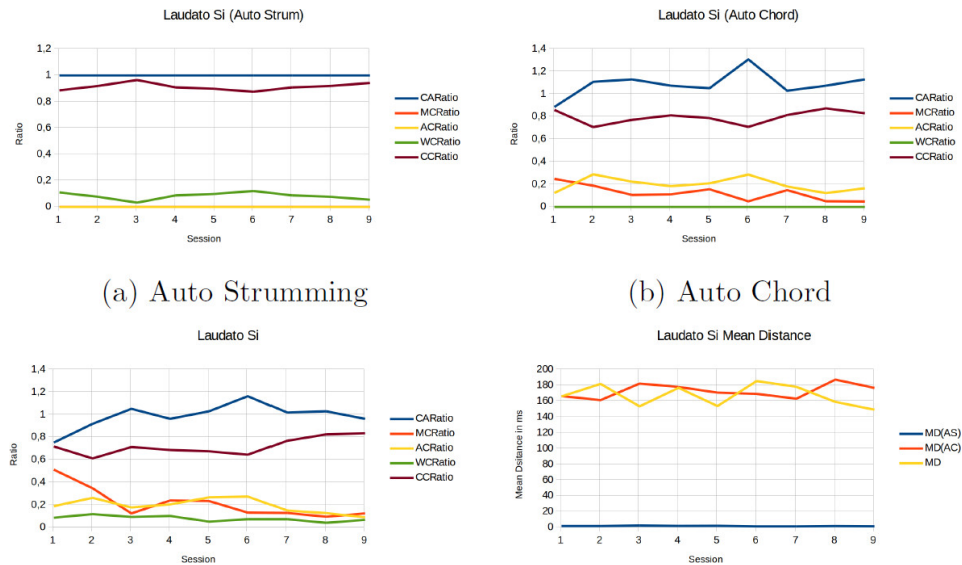
Figure 2.24: Results for the song "LaudatoSi".

Usability Test and User Experience Test

For the usability test the subject was able to fulfill all the given tasks.
However as an observer the following problems were noticed: (i) buttons to
change modes were not clear to the user, (ii) a change of hand positions was
observed when using auto strumming mode and the other modes, (iii) the
hand moved on the sledge while sliding (iv) the user changed fingers while
strumming.

When the user was asked to express his feelings about the project he stated:
"I'm happy that I can finally learn to play the guitar. It is still difficult
to play the songs, especially the strumming. But as with a traditional
instrument, it needs a lot of practice to become an expert. With the auto
strumming activated I can play already a few songs and I even gave a
small performance to my colleagues. When Roberto (his supervisor) was
accompanying me with the piano I felt like playing in a band and it made
me really proud when my colleagues were listening to me playing guitar.
My goal is to practice a lot so I can play at our workshop's anniversary
next year. It is still a long way but I'm eager to practice."

Figure 2.25: Results for the song "Hallelujah".

2.4.5 Discussion

When looking at figures 2.23, 2.24 and 2.25 for each song (with auto strumming activated) we can see that the chord amount ratio (blue) stays one. This is because auto-strumming is activated and therefore the amount of played chords equals the amount of supposed chords.

We can also see a slight improvement of correct chords over the nine weeks for all three songs in both auto strumming and auto fretting modes. The performance with no support is as expected still much lower compared to the performances with auto modes activated. However, even here we can see an improvement from the first session to the last session. It is interesting to see that the performance of the most complex song increased the most in the case of no support.

The mean distance from the correct chords to the original chords stayed around 180ms for all three songs which indicates that the strumming is still random. The maximum distance from a played chord to the original chord at a speed of 80 BPMs and a down strum for every beat is 375 ms, and therefore a random strumming would average between 0ms and 375ms which is 187ms.

The usability test suggests a few improvements for upcoming prototypes.

For instance the functions of the software buttons seem to be unclear and should have a more clear design. A small label could for instance enhance the understandability of the graphical user interface. Another very important observation is that the user applies different hand positions when playing the interface in different modes. When auto strumming is activated he rests his left hand on the edge of the sledge, so just his fingers are on the sledge, but his palm is in front of the sledge and his thumb is being used as a "brake". However, with this hand position he is not able to reach the wheel with his fingers and if he has to strum he has to move his hand to another position. It was also observed that his hand moves around the sledge when sliding movements are executed. This leads to the fact that he has to use different fingers for the strumming depending on his hand position. A broader wheel as well as some kind of hand guidance could be implemented to avoid those problems. To see if those changes would aid the user and improve his/her performance additional testing would be needed.

Although the performance measured quantitatively in general was poor (e.g. the strumming is random), the user shows signs of satisfaction when performing. The subject might not acquire too high performance skills when compared with guitarist without disabilities, but he/she might receive the therapeutic benefits of playing a musical instrument.

Two years after giving the prototype do the subject, we conducted him/her and asked him on his experience on the prototype. He replied that he still uses but not really regularly. According to his therapist, he still prefers to play with auto-strumming turned on, and he stills enjoys playing songs with it.

2.5   Conclusions

In this chapter we presented the implementation and evaluation of 3 different prototypes designed for 3 distinct cases of people with limited upper limb function.

The percussive midi controller constructed for a person with spinal cord injury proved to have a big positive impact in his quality of life. This user was already a musician before the injury. Before providing him with the constructed prototype he wouldn't manage to play music with any instrument. Two years afterwards he is an active music composer, and he plays music in a daily basis, using the constructed prototype to improvise alone or with friends and compose music.

The guitar controller for a person with limited sensorimotor functions allowed our subject to easily change chords with the left hand while strumming with the right hand, as she would do in a normal guitar. The prototype although still lacks important improvements that would make it more usable.

The guitar controller for a person with cerebral palsy had a therapeutic impact in his life. Although we did not perform any additional evaluation (apart from the evaluation performed 2 months after with experience on the prototype), when we asked his therapists, they responded that he/she still enjoys playing with it, although he might have only slightly improved his performance.

The importance of all mentioned studies is that, using arduino microcontroller and low cost sensors and materials, we were able to construct musical interfaces that allowed people with limited upper limb functioning to play music, and enjoy its psychological benefits. As instructions of how to implement them are available in this thesis (and with more technical details and source code, online), they could be used as examples of how other people could construct adaptive digital music instruments for people with disabilities in their community.

Gaze-Controlled Music Interfaces

3.1 Introduction

In more severe cases of motor disabilities, such as people with locked-in syndrome, none of the solutions mentioned in previous chapters is appropriate. Locked-in syndrome (LIS) is a condition in which a patient is conscious but not able to move or communicate verbally due to complete paralysis of nearly all voluntary muscles in the body except the muscles which control the eyes. In such cases a gaze or brain controlled interface might be only mean of communication. In this chapter we present our research on gaze controlled musical interfaces.

In order to clearly see an object of interest, we have to place it on our fovea Yarbus (1967). The fovea is a high-acuity region of the retina that covers approximately one degree of visual arc, which is a bit less than the width of the thumb of an extended arm. Switching from one point of interest to another is achieved through abrupt, fast eye movements, called saccades. Saccades last between 30 and 120 ms (Krauzlis (2005)). The time in which the gaze maintains at a single location, is a called fixation. A fixation lasts at least 100 to 200 ms. During fixations, the eyes make small, jittery motions of three types: tremor, drift and microsaccades (Martinez-Conde (2006)). Tremors are movements of high frequency ($\approx$80Hz) and low amplitude ($\approx 0.0024°$ visual angle). Microsaccades are fast movements that occur 3-4 times every second. Drifts are slow motions that occur between microsaccades. Drifts and microsaccades have an amplitude of 0.03 to 2 °

of visual angle. In gaze-controlled applications, in which the fixations are usually mapped to a pointer, smoothing algorithms eliminate the fixation micromovements and the noise introduced by the video based tracking of the pupil.

In eye-tracking-based (gaze-controlled) applications, gaze data might be used alone or in combination with other input methods, such as head, limb or breath-controlled buttons. Blinking (closing both eyes) or winking (closing just one eye) might also be used as input. In this case, usually the gaze coordinates are used for pointing, and any other input is used for triggering actions. In case the gaze input is used alone, as the eye movements are often non-intentional, gaze information must be interpreted carefully to avoid unwanted responses to user actions. This is described as the "Midas Touch" problem. The most common gaze selection methods that intend to handle the Midas touch problem are: (i) Dwell time introduced by Jacob (1991) and (ii) screen button introduced by Ohno (1998). In the case of the dwell time method, when a fixation lasts for more than a given time period (typically about 1 second), a selection is made. In the case of screen button method, each target is separated in the command name area and the selection area. Selections are made only when a fixation is detected in the selection area. An extension of the screen button method is the pEYE method introduced by Huckauf and Urbina (2008), in which the "slices" of the "pEYE" are screen buttons. The command name areas of the buttons are placed at the interior of the pie, and the selection areas are placed at the perimeter. Other selection methods that also handle the problem of the noisy gaze data involve magnification methods, like ZoomNavigator proposed by Hegner and Skovsgaard (2008)).

An extensive review of eye-controlled music performance systems was recently made by Hornof (2014). Some of these installations do not aim to resemble traditional musical instruments: they could be described as sonifications of eye movements and they are not designed for playing melodies. Here we will only refer to the approaches that provide the possibility of playing separate notes. Duet for eyes[1] was a performance including performers with and without disabilities. The Grid software [2], using dwell-time selection method and controlled by a Tobii eye tracker [3], was used to trigger preselected sounds. In a more recent project, called "eye play

---

[1] http://illustriouscompany.co.uk/performance/duet-eyes-eyejamming-eyebodyweaving, last accessed on 10/12/2015

[2] http://sensorysoftware.com/, last accessed 2015-4-12

[3] http://www.tobii.com/, last accessed on 10/12/2015

the piano" [4], by the University of Tsukuba and FOVE eye tracking virtual reality headset[5], people with disabilities were able to trigger notes or chords of a piano, assigned to buttons on the screen. Blinking was used as a selection method. In the mentioned gaze-controlled setups dwell-time and blinking selection methods are used for triggering musical events. However, dwell-time does not allow triggering events in tempo, while blinking selection requires two actions in order to trigger a single event: (i) focusing on a target and (ii) blinking. When compared to the screen-button selection methods, the blinking method provides a slower interaction time. Moreover, none of the mentioned systems allow the control of more expressive musical features, such as loudness.

3.2 The EyeHarp DMI

We present the EyeHarp, a gaze controlled musical interfaces that aims to allow a rich and expressive gaze-controlled music performance. The EyeHarp allows the user to control pitch, timing and dynamics of a melody, as well as chords and arpeggios in a performance. The EyeHarp interface consists of two layers: the Step Sequencer layer and the Melody layer. In the Step Sequencer layer chords and arpeggios can be constructed and in the melody layer these can be controlled and a melody can be played. The number of available note buttons can be adapted according to the accuracy of the eye tracker and the expertise of the performer. The user can switch between the two layers through a dwell-time activated button.

The EyeHarp is implemented using openFrameworks open source C++ toolkit[6]. It has a built-in analog synthesizer and it also works as a midi device, controlling any external software synthesizer. The EyeHarp is currently an open-source software[7] that runs in windows 7 or later operating systems. Currently two commercial eye-trackers are supported: the Eyetribe[8] and Tobii PCEye[9]. The non-commercial open-source ITU Gazetracker[10] is also supported. In all three cases the EyeHarp receives through a server the raw gaze data. Fixation detection and smoothing algorithms

---

[4]http://eyeplaythepiano.com/en/, last accessed on 10/12/2015
[5]http://www.getfove.com/, last accessed on 10/12/2015
[6]http://www.openframeworks.cc/, last accessed on 14/12/2015
[7]Source code and binaries available at https://github.com/zackbam/TheEyeHarp
[8]https://theeyetribe.com/, Last accessed on 17/12/2015
[9]http://www.tobiidynavox.com/pceye-go/, Last accessed on 17/12/2015
[10]http://sourceforge.net/projects/gazetrackinglib/, Last accessed on 17/12/2015

are incorporated in the EyeHarp software. This allows a consistent behavior of the system when different eye trackers are used.

The interface is diatonic and by default tuned to the C major scale, although it can be tuned to any possible scale. Only the basic functionality of the EyeHarp interface will be described here. A more detailed overview of the more advanced features of the interface was presented by Vamvakousis and Ramirez (2011).

3.2.1   Gaze data processing

Video-based eye trackers normally track the pupil of the user and the reflections of a number of infrared light-sources on the cornea in a constant frame rate. Even when the user is looking at the same point for many consecutive frames, the received raw gaze coordinates will be spread in a Gaussian-like distribution around a point. Smoothing algorithms are typically applied before using the raw gaze data in gaze-pointing tasks. A smoothing algorithm has the form of:

> **while** *Receiving Gaze Data* **do**
> $\quad smooth.x \leftarrow SF \cdot smooth.x + (1 - SF) \cdot raw.x;$
> $\quad smooth.y \leftarrow SF \cdot smooth.y + (1 - SF) \cdot raw.y;$

SF is the smoothing factor (greater than 0 and smaller than 1), smooth.x and smooth.y are the gaze coordinates on the screen and raw.x, raw.y are the raw gaze coordinates per frame.

The above algorithm introduces a latency proportional to the distance of the two consecutive fixation points. For that reason, typically, when a saccade is detected, the smoothed coordinates maintain the last value of the previous fixation until a new fixation is detected and the are assigned to the value of the new fixation point.

Although the accompanying software of both the Eyetribe and the Tobii PC-eye eye trackers provide the smoothed gaze coordinates, they were judged insufficient for the purpose of playing melodies in the EyeHarp pie: when the user was playing two adjacent notes, no saccadic movement was detected. As a result the pointer was moving smoothly from the one fixation point to the other. The saccadic movement was successfully detected only when the consecutive focus points were distant enough. This led in poor temporal control when playing adjacent notes. Moreover the interface did

not behave with consistency when using different eye trackers. The smoothing algorithm of the EyeTribe and the Tobii PCeye is appropriate only for gaze-controlled applications in which the temporal accuracy is not a crucial factor. In order to achieve a consistent system in terms of temporal control, the smoothing algorithm was re-implemented in the EyeHarp interface, allowing the adjustment of its parameters in such a was that fine temporal control in achieved for small or large saccadic movements. The evaluation of the implemented fixation detection algorithms will be presented in sections 3.3 and 3.4.

3.2.2   The Step Sequencer Layer

Figure 3.1 shows the Step Sequencer layer. A step sequencer is an interface for constructing loops. It consists of a grid of buttons where the vertical dimension of the grid corresponds to pitch and the horizontal dimension corresponds to the temporal position in the loop. At the beginning of the loop, the selected notes of the first column sound simultaneously, followed by the selected notes of the second columns, and so on. After the notes of the last column are played, the loop starts from the beginning. The time interval between the activation of two consecutive columns is constant and depends on the set tempo.

In order to select a button of the step sequencer, dwell-time selection is applied (the default dwell time value of the EyeHarp interface is 700 ms). The buttons are circles with a small focus point in the center. The focus points help the user focus at the center of each button. The Step Sequencer layer has two methods for compensating noisy gaze-data. In the first method the gaze point appears on the screen along with additional fixation points at the perimeter of the buttons. This helps the user correct the offset caused by poor tracking. In the second case, when a fixation is detected and the dwell time period is reached, the buttons of the step sequencer that are placed within a square region -centered at the fixated point and covering the 20% of the sequencer area- are magnified by a factor of 2. The user can then select one of the magnified buttons. By looking outside the magnified area, all buttons return to their normal size and position. Figures 2 and 3 demonstrates the two described methods. Note that in case of the magnification method, as the buttons expand, they might come out the screen. In that case all magnified buttons smoothly move up or down in order to appear inside the screen.

A number of control buttons (e.g. for changing the meter or tempo, clear-
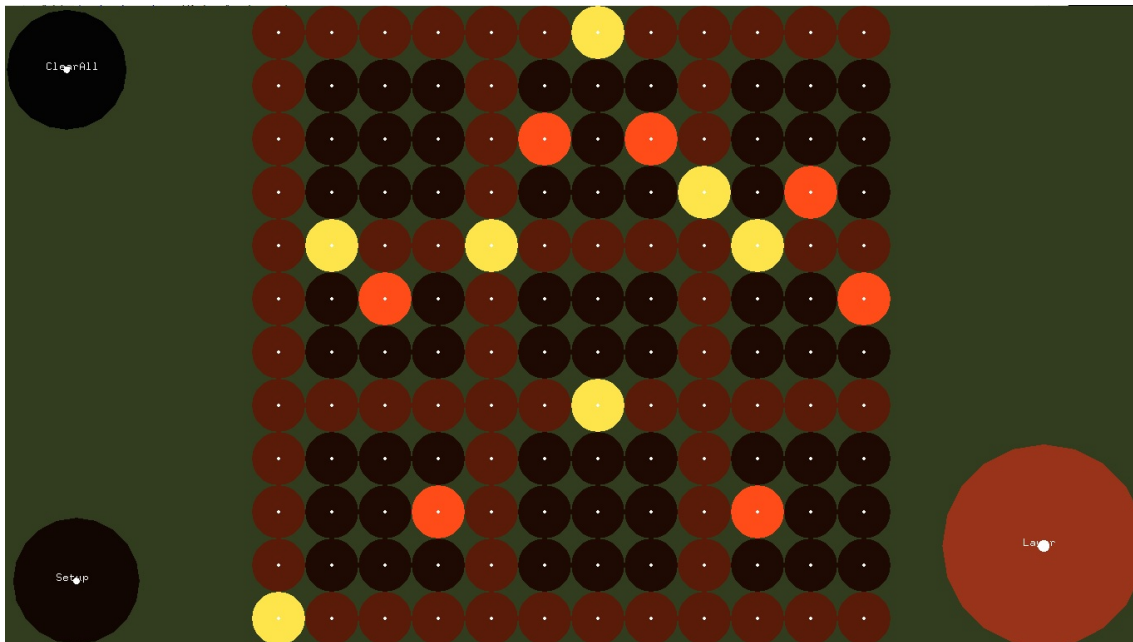
Figure 3.1: The Step Sequencer Layer. In this layer the user can construct arpeggios and chords which are controlled in the Melody Layer. Buttons in the same row correspond to notes with same pitch, while buttons in the same column corrspond to simultaneous notes. If the selected chord in the melody layer is C major, buttons from bottom to top correspond to notes C4, D4, E4, etc. Notes are trigered from let to right, starting with the left most column. Dwell-time selection method is used, i.e. users focus at each button for about 700 ms in order to select or release a button..

ing the selected notes, switching between layers) are provided and may be selected using dwell-time. (see figure 3). The note that corresponds to bottom row of the EyeHarp's Step Sequencer corresponds to the base note of the selected chord in the melody later. The notes corresponding to the other rows in the step sequencer are mapped to the consecutive notes.

3.2.3   The Melody Layer

The Melody layer (Figure 4) is based on pie menus. A pie menu is made of several "pie slices". Each slice consists of an inactive area in the center and the selection area at the perimeter of the circle. The idea of using pie menus in gaze interaction was introduced by Huckauf and Urbina (2008) in a typing and a desktop-navigation interface. The idea of the pEYE layout is

appealing for playing melodies because clicking is not necessary for making a selection. Once the pointer enters in the selection area at the perimeter of the pie, a command is triggered. The slices of the pie menu of the Melody layer can be thought as screen buttons (as introduced by Ohno (1998)). The command name area is a number for each note and a Latin number for each chord. The selection area is placed at the perimeter of the pie. At the center of the selection area of each note, a focus point appears. Optionally as shown in figure 5, multiple focus points appear in the selection area of each slice. Outer focus points correspond to high loudness, while inner points correspond to lower loudness.

If the set scale is C major, c in the 4th octave is placed at 180°. The scale then goes up counterclockwise. As a default option the pie comes with 14 slices, but the number of slices can be adapted though the setup menu. If the setup button is pressed in the melody layer, a number of configuration buttons appear as shown in figure 5. Two repeat buttons on the left can be used for adjusting the number of notes in the pie. Through four radio buttons on the top the user can select between three preset sounds of the EyeHarp internal synthesizer, or select the midi out option. In that case the interface is sending midi messages to an external synthesizer through the LoopBe virtual midi port[11].

If the "chords" buttons is active, the last 6 notes of the pie are replaced by 6 chords. These buttons control the harmony of the arpeggio constructed in the Step Sequencer layer. In order to play a note or change the chord, the user can either look directly at the selection area of the note/chord or in case there is a big distance between two consecutive notes, the command name area might be looked at before focusing on the selection area. Focusing on close targets is more efficient than focusing on distant ones. Fitt's law also applies to gaze interaction as shown by Miniotas (2000). In order to release a note, the user has to look at any place outside the pie. For that reason some fixation points are placed outside the pie. When a fixation is detected at the selection area of a note the note sounds and a button appears at the center of the pie. This allows the user to repeat the same note twice. If a fixation is detected inside the button's area, the same note sounds again. If a fixation is detected at any other point apart from the selection area of the pie and this button at the center, the "repeat" button disappears.

A fixation detection algorithm is applied on top of the smoothed gaze coordinates. A selection is made only if two consecutive gaze points are detected

---

[11]http://www.nerds.de/en/loopbe1.html, last accessed on 17/12/2015

inside the selection area of the same note and if the distance between them is less than 2 pixels.

3.3 Temporal Control in Gaze Interaction

Probably the most important characteristic of music is rhythm. Previous research for finger tapping has shown that people can accurately tap out rhythms with their fingers as fast as one tap every 100 ms, and that people tend to tap a few tens of milliseconds before the beat but that this negative mean asynchrony decreases and disappears with musicians. An extensive review of the finger tapping synchronization studies was performed by Repp (2005).

The research in eye movement sensorimotor synchronization though is limited. In a study performed by Hornof and Vessey (2011) the temporal accuracy of two fixation and two saccade-based methods was evaluated. Participants moved their eyes back and forth between two small squares on a computer display to play handclap sounds to attempt to match a rhythm of woodblock sounds, in three different tempos: 60 beats per minute (bps), 120 bps and 240 bps. The two small squares were centered on the display and separated by 12° of horizontal visual angle. A vertical midline separated the two squares. The two fixation-based trigger methods were: (i) dispersion-based, in which a fixation is detected when the raw gaze date are 0.5° visual angle close for more than 100 ms and (ii) velocity-based, in which a fixation is detected when the movement of the gazepoints across the display holds below 20° per second for 100 ms. The two saccade-based methods were the (a) maximum velocity detection-method, in which the tap was triggered by the first gaze sample after maximum velocity of the saccade, and (b) the midline crossed condition, in which the tap was triggered by the first sample across the midline drawn on the display.

The typical metric to describe temporal accuracy is the asynchrony, measured in milliseconds. Consistent with Repp (2005), if a note is played earlier there is negative asynchrony, while if it is played later, there is positive asynchrony. The results of the study indicate that fixation-based gaze-selection algorithms provide more accurate rhythmic and timing control than saccade-based gaze-selections algorithms, and that people have a fundamental performance limitation for tapping out an eye-controlled rhythm somewhere between two and four beats per second. The saccade-based methods resulted in an average negative asynchrony of around -50 ms across

subjects when playing at a beat of 60 bps. For the same condition, the fixation-based methods resulted in an asynchrony of less than -8 ms. It is important to point out that tested fixation-based algorithms introduced a latency of at least 100 ms from the moment a saccade starts, while in the case of the saccade-based methods, the introduced latency is less than 20 ms.

The study described in this section was presented at the Proceedings of the 12th International Conference on NIME (Vamvakousis and Ramirez (2012)).

Evaluating the temporal accuracy of the Melody Layer

Materials & Methods   Utilizing the dispersion-based fixation detection algorithms we conducted an experiment studying the temporal accuracy in the melody Layer of the EyeHarp DMI (Vamvakousis and Ramirez (2012).

The open source desk-mounted binocular ITU gaze tracker was used in this study, running at 30 Hz. It consists of a modified playstation 3 "Eye" camera and a pair of infrared led arrays (see figure 3.2) . Although the initiative of the ITU gaze tracker is now discontinued by its founders, the source code can still be found online[12] .

Ten healthy subject, with at least 5 years of musical training, were asked to perform 2 different tasks in the Melody Layer. A metronome set at 60 bps was sounding during both tasks. Subjects were asked to play in-tempo one note per 2 beats. Afterwards they were asked to perform the same task using a computer keyboard, with number '1' assigned to the first note of the scale, number '2' to the second and so on.

The first task consisted of playing an ascending scale. The task was repeated 3 times. Before each repetition they practiced for about 2 minutes. The purpose of that was to study whether there is any improvement in the temporal accuracy with practice. Note that this task consisted of playing only adjacent notes in the pEYE interface .For reasons of consistency in both cases the frame-rate was set to 30Hz.

The second task consisted of playing an octave interval for 40 times. As in task 1, it was asked to play one note every to beats.

Results & Discussion   Figure 3.3 shows the asynchronies of all 10 subjects for all 8 notes of a scale and all 3 repetitions of the first (playing scale)

---

[12]https://sourceforge.net/projects/gazetrackinglib/ last accessed on 19 of May, 2016
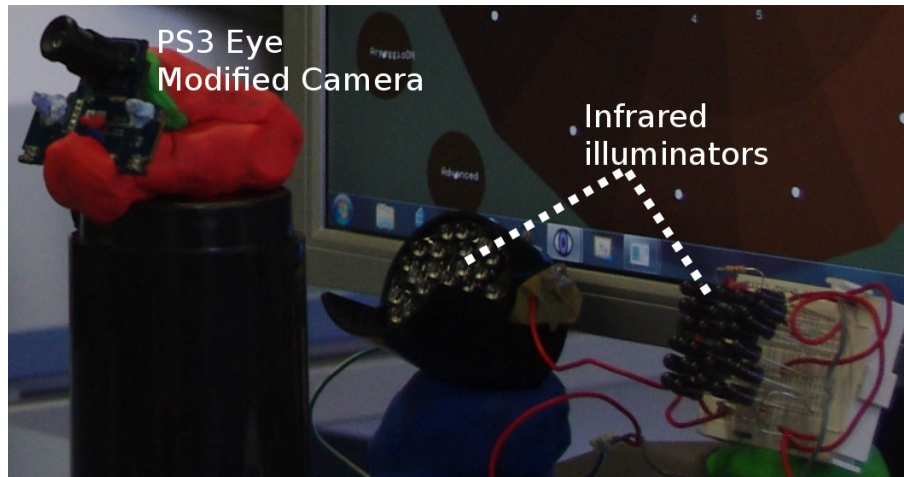
Figure 3.2: The PS3 Eye camera was modified by (i) replacing its lens, (ii) adding an infrared-pass filter. Two arrays of infrared leds were placed to the left and right of the camera. Both the infrared-led arrays and the camera when then placed under the computer screen. The ITU gaze tracker open source software was then used for calibrating and acquiring the gaze coordinates. Instructions of how to transforms the PS3 camera into an eye-tracking device were found online. Although the source code of the project is still online, the support forum is currently down, as the founders if the initiative currently offer a commercial low cost eye tracking device called Eyetribe.

task. The average asynchrony across all subjects and repetitions was -95 ms and the standard deviation of the 10 participants means 33.5 ms. The mean asynchrony and standard deviation for the first repetition was -64 ms and 410 ms respectively, for the second -124 ms and 252 ms, and for the third -96 ms and 211 ms. The fact that the standard deviation is gradually falling, suggests that there is an improvement with practice. Nevertheless a more prolonged time of practice is required to confirm this hypothesis.

Figure 3.4 shows the same results for the case of the task where the keyboard was used as input. When using the keyboard as input t average asynchrony is positive. The standard deviation gradually falls. As expected, the standard deviation values are much lower when the keyboard was used as input method.

Figure 3.5 shows the mean asynchrony across all subjects for the 1st and 2nd task. The mean asynchrony for the first task is -46 ms and for the 2nd task -95 ms. The first task consisted of playing a ascending scale. In this
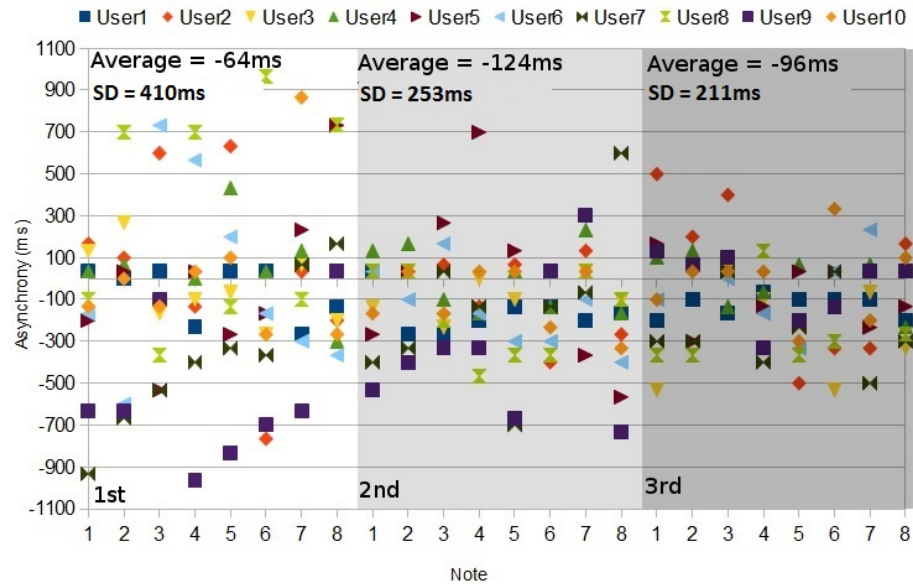
Figure 3.3: The asynchrony of each user for all 3 trials of playing an ascending scale in the Melody Layer. The horizontal axes corresponds to the number of the note (instance) in each task.

case all notes are adjacent (about 5° of visual angle apart). The second task consisted of playing notes placed diametrically opposed (about 18° of visual angle apart). This suggests that the negative asynchrony increases with distance.

3.4 A new Fixation Detection Algorithm for Improved Temporal Accuracy of Distant Targets

3.4.1 Introduction

As already mentioned, Hornof and Vessey (2011), in a study involving 12 musicians, compared the rhythmical precision of producing clapping sounds by looking back and forth between two small squares separated by 12° of visual angle. In that study the two factors were trigger method and tempo. The results suggest that fixation-based eye-control algorithms provide better timing control than saccade-based algorithms. The fixation-based al-
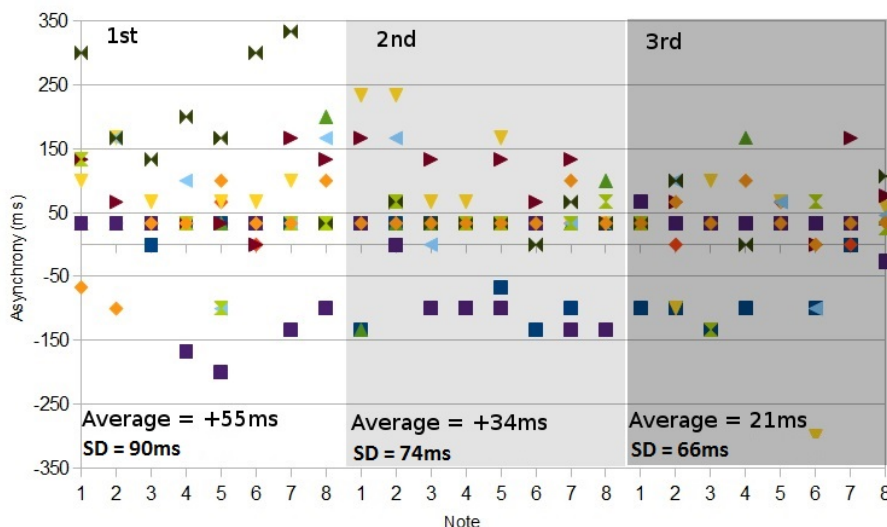
Figure 3.4: Same as figure 3.3, but for the case that the keyboard was used as input.

gorithms tested in that study imposed a fixation detection latency of 100 ms. The saccade-based algorithms triggered clapping sounds even before a fixation would have started, as the event was triggered when a saccade was detected and before a new fixation started.

The results of our study described in 3.3 suggest that asynchrony increases with distance. This is conforming to the fact that the duration of a saccadic movement is a linear function of its amplitude. Collewijn et al. (1988a) in a study involving 4 healthy subjects concluded that the duration of saccades of amplitude less than 50° of visual angle can be well approximated by the following linear function:

$$durationH = 2.7 \cdot amplitude + 23ms \qquad (3.1)$$

Collewijn et al. (1988b) in another study concluded that vertical saccades of amplitude less than 30° of visual angle can be well approximated by the following linear function:

$$durationV = 3.3 \cdot amplitude + 31ms \qquad (3.2)$$

The most commonly used algorithms for detecting fixations are grouped into two categories: (i) The velocity-based algorithms. In that case if the
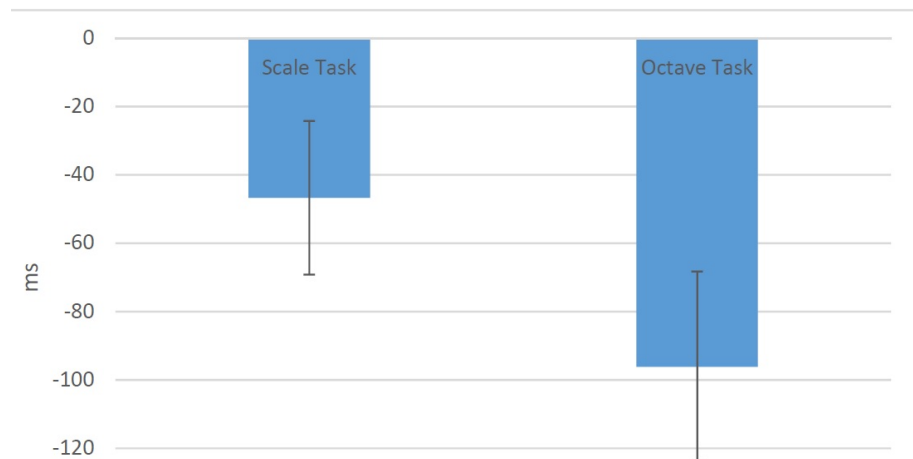
Figure 3.5: Mean asynchrony for the first task and the second task. The first task consisted of playing a ascending scale. In this case all notes are adjacent (about 5° of visual angle apart). The second task consisted of playing notes placed Diametrically Opposed (about 18° of visual angle apart). The standard error of the 10 participant means is also shown.

distance between two consecutive gaze points is smaller than a threshold, a fixation is reported, otherwise a saccade is reported. (ii) The dispersion-based algorithms. In that case, the centroid of N frames is computed. If the geometric distance of all N frames from the centroid is within a threshold, a fixation is reported, otherwise a saccade is reported. Variations of the described velocity and dispersion-based fixation detection algorithms are commonly used in gaze interaction applications (Shic et al. (2008); San Agustin (2010); Larsson (2010); Salvucci and Goldberg (2000); Kumar et al. (2007)).

Normally velocity-based fixation-detection algorithms impose a small detection latency, corresponding to just one frame. If we suppose that the frame-rate is 60 Hz, this corresponds to 16.6 ms. On the other hand, dispersion-based algorithms commonly impose a latency of 100 ms.

Driven by the results of the study presented in 3.3, for the purposes of musical gaze controlled applications, in which temporal accuracy is crucial, we propose and evaluate a dispersion-based fixation-detection algorithm whose fixation detection latency depends on the saccadic amplitude. This is achieved by using different size of time window used for computing the

dispersion value, depending on the saccadic amplitude.

3.4.2 Methods

The experiment was a 3X3 within-subjects design. The two factors were (i) fixation-detection algorithm and (ii) distance between targets.

Eye tracking data were collected by the eyetribe binocular desk-mounted 60 Hz eye tracker. The eye tracker was connected to a laptop equipped with a 15.6 inches screen with resolution 1366X768, an Intel i5 M460 2.53 GHz dual core CPU running windows 10 operating system. Asio4All Universal ASIO Driver was used with buffer size 64 and sample-rate 44100 Hz offering 1.5 ms of audio latency. A chinrest maintained an eye-to-screen distance of 60 cm.

Seven male adults with moderate to professional level in playing a musical instrument were recruited from the Music Technology Group research center of Universitat Pompeu Fabra. Each participated for about 20 minutes and completed nine two-minutes blocks. Each block was performed with one of the three implemented fixation detection algorithms. The ordering of the blocks was randomized, and counterbalanced across participants. The first six blocks were to practice with all algorithms (2 blocks for each algorithm) and we report on the average asynchrony of the last three blocks.

Participants moved their eyes back and forth between two circles on a computer display to play synthesized sinusoidal sounds, to attempt to match the beat given by a high pitch sinusoidal sound. All sounds had an attack phase lasting 1 ms followed by a release phase lasting 249 ms. The screen was divided horizontally in 4 equal parts. The position of the left circle was at the center of the left part. The position of the right circle was at the center of one of the remaining parts. In particular, at the beginning of the experiment, it was at the middle of the second from part from the left. The moment the user completed a left-right-left eye movement, the position of the right circle moved to the center of the second part from the left. After another left-right-left eye movement the right circle moved to the middle of the third part from the left. This resulted in three possible saccade amplitudes of visual angles: $9.2°$, $17.9°$ and $25.8°$. One repetition consisted of performing all three saccadic amplitudes.

The following visual feedback were given to the participants during the experiment: If a note was performed earlier, the color of the button corresponding to it appeared blue, while if it was performed later it appeared

red. The brightness of blue or red depended on the degree of asynchrony. The objective of the participants was to maintain the color of the buttons dark.

The three implemented and studied fixation-detection algorithms were:

(i) Velocity-based (VEL). If the distance between two consecutive gaze points is smaller than 1.2° of visual angle (=50 pixels), a fixation is reported, otherwise a saccade is reported. In that case, fixation-detection latency is 16.6 ms.

(ii) Dispersion-based (DISP). The centroid of the last the gaze points corresponding to 100 ms, is computed. If the distance of all these points is smaller than 1.2° of visual angle (=50 pixels), a fixation is reported, otherwise a saccade is reported. In that case the fixation-detection latency is 100 ms.

(iii) Dispersion-based of varying time window (VARDISP). In that case the number of gaze points contributing in the calculation of the dispersion is dependent on the amplitude of the saccadic movement, in such a way that the time required to report a fixation from the moment a saccade starts is the same for all amplitudes of saccade movement. Equation 3.1 was used to compute the total fixation-detection amplitude. In specific, the fixation-detection time was computed using the following formula:

$$FixationTime = 140 - (2.7 \cdot VisualAngle + 23)(ms) \qquad (3.3)$$

The value of 140 ms was selected having in mind that the algorithms should introduce the same amount of latency as the dispersion-based algorithm for the low-amplitude saccades performed in the experiment. According to this formula, the fixation-detection latency introduced by the algorithm reduces linearly with the amplitude of the saccadic movement. Fixation time value is computed for every new gaze point received from the eye-tracker. The Visual angle corresponds to the angular distance of the current gaze point from the smoothed gaze coordinates. The implemented VARDISP algorithm introduces 100, 66.4 and 49.8 ms fixation-detection latency for each of the 9.2°, 17.9° and 25.8° visual angles respectively.

The overall variance of asynchrony across all 7 subjects, algorithms and saccade amplitudes was computed and asynchronies with absolute values more that two times the variance were excluded as outliers. Additionally, all gaze

data were stored in a cylindrical buffer along with a timestamp. The average frame-rate of the last 20 gaze-data received from the eye-tracker server was computed for each recorded instance of the experiment. If this value was more than 32 Hz, the instance was excluded from the analysis. Any frame-rate lower than 60Hz would be a result of lost gaze frames due to bad tracking or inefficient connection of the eye-tracker. We assumed that such a behavior would make the system less responsive, and this might affect the recorded asynchronies. After applying both data-exclusion criteria, in total 9% of all samples were excluded as outliers. For each saccadic amplitude both the left-right and the right-left eye-movement were taken into account.

At the end of the session, subject were asked which was the algorithm that in their opinion would allow better temporal accuracy.

3.4.3 Results

Figure 3.6 shows the average asynchrony and standard error across all subjects, for all three tested fixation-detection algorithms.

An ANOVA was performed for all 3 saccadic amplitudes comparing the VARDISP algorithm with VEL algorithm and DISP algorithm. In the case of the 9.2° saccadic amplitude, no statistically significant differences between the mean asynchronies were detected (VARDISP vs VEL p-value = 0.27, VARDISP vs DISP p-value = 0.93). In the case of the 17.9° saccadic amplitude, statistically significant differences were found between mean asynchronies of the VARDISP and VEL algorithms (p-value = 0.05), while no statistically significant differences were found between the mean asynchronies of VARDISP and DISP (p-value = 0.87). In the case of the 25.9° saccadic amplitude, statistically significant differences were found between mean asynchronies of the VARDISP and DISP algorithms (p-value = 0.003), while no statistically significant differences were found between the mean asynchronies of VARDISP and VEL (p-value = 0.29).

All participants declared that they preferred the VEL algorithm. Most subject agreed that although indeed they tended to play earlier when using this algorithm, they liked the fact that the algorithm was more responsive.

3.4.4 Discussion

The above results make sense if we consider that the behavior the VARDISP algorithm is similar to that of the VEL algorithm for short saccadic movements and similar to the DISP algorithm for long saccadic movements. In
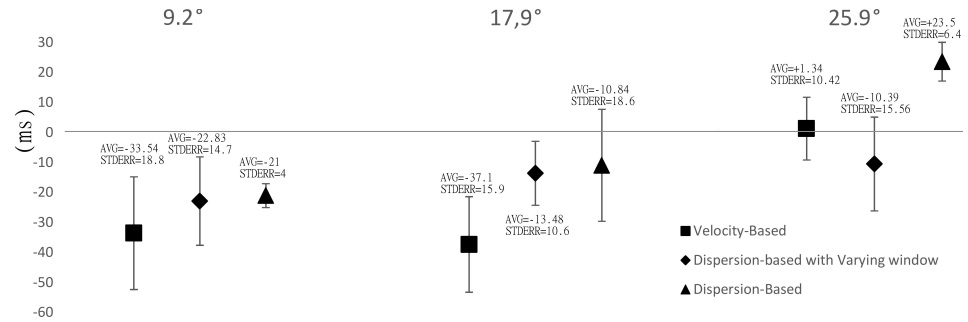
Figure 3.6: Average asynchrony and standard error across all subjects, for all three tested fixation-detection algorithms.

the case of the low-amplitude saccades (9.2° and 17.9° of visual angle) the proposed VARDISP algorithm has mean asynchrony similar to the DISP algorithm and both algorithms perform better than the VEL algorithm. An explanation for that could be that saccadic movements of small amplitude are short in time. Introducing latency before triggering an action reduces the absolute value of the asynchrony. Similar results were reported by Hornof and Vessey (2011), were in the case of 12° of visual angle, the dispersion-based algorithm (introducing 100 ms of detection latency) performed better saccade-detection algorithms (introducing 16.6 ms of detection latency).

Nevertheless, in the case of 25.9° of visual angle saccadic movements, the DISP algorithm's mean asynchrony is +23.5 ms. This can be explained by the fact that long saccadic movements last longer. In that case, reducing the detection latency improves the mean temporal accuracy.

The VEL algorithm gives the best mean temporal accuracy for the 25.9° of visual angle saccadic movement, when compared to the other two algorithms. Nevertheless, in the case of 9.2° and 17.9° of visual angle, it gives the highest negative asynchrony. Although the DISPVAR, in the case of the 25.9° of visual angle, introduces 34 ms more detection latency than the VEL algorithm, its mean asynchrony is 11 ms less than the mean asynchrony of the VEL algorithm. Although this difference is not statistically significant (p-value=0.29), a possible explanation might be that the subjects might expect higher latency (as in the case of smaller saccade amplitudes) and thus might choose to start the saccade earlier.

Although introducing varying fixation-detection latency seems to improve

the overall mean temporal over traditional fixation-detection algorithms - that introduces a steady amount of latency-, it is surprising that subjects expressed a preference over the velocity-based algorithm. The question that rises is whether through long-term practice, a low-latency fixation-detection algorithm would give better temporal accuracy than any sophisticated algorithm that gives better results for novice EyeHarp players. Through practice, the user might train himself to start short saccadic movements later than short ones.

3.5   Evaluation of the EyeHarp DMI

The studied described in this section has been published in Frontiers in Psychology (Vamvakousis and Ramirez (2016)).

O'Modhrain (2011) proposed that a DMI can be evaluated from the perspective of (i) the audience, (ii) the performer, (iii) the designer and (iv) the manufacturer. In this evaluation process we evaluate the proposed interface from the perspective of the audience and the performer.

3.5.1   Audience Perspective

Transparency describes the level to which a performer or spectator can understand the relationship between the input (gesture) and output (sound). According to Hunt et al. (2002) and Arfib et al. (2005), an instrument's capability for expressive performance is positively correlated to its degree of transparency, i.e. how clear is the mapping between the gestures of the performer and the sound produced by the instrument. Unlike traditional musical instruments, in DMIs the way the performer's gestures produce sound might not be physically evident to the audience. Schloss (2002) suggests that the lack of an obvious connection between cause and effect dramatically affects the way a performance is perceived by the audience. According to Schloss, providing visual cues that aim to reestablish the connection between cause and effect is a key component in making a DMI performance convincing and effective.

Reeves et al. (2005) proposed an evaluation of DMIs based on audiences perception of the relationship between input manipulations and audio output. They characterize a performance with low input and output comprehension as "secretive", one with low input and high output comprehension as "magical", one with high input and low output as "suspenseful", and one with high input and output as "expressive". Barbosa and Calegario

(2012) extended Reeves's classification and proposed 5 different aspects to be considered in building the "interaction model" of a DMI: (i) The cause comprehension refers to how clear the available input gestures are. (ii) The effect comprehension refers to how clear the controlled parameters are. (iii) The mapping comprehension refers to how clear is the relation between user's actions and the resulting sound. (iv) The intention comprehension refers to what degree the system allows the user to express his musical intentions. (v) The error comprehension refers to whether the possible errors in the performance were noticeable.

A concert was organized at the concert hall of Universitat Pompeu Fabra. The performer had been practicing the EyeHarp for a period of 10 weeks, playing 3 times a week. Every practice session lasted for approximately 20 minutes. The concert consisted of two parts. In the first part the EyeHarp player performed a piece composed by him for EyeHarp solo performance and in the second he performed along with two guitar players and a flute player in a jam session. One of the eyes of the performer was shown at the center of the screen and the coordinates of his gaze were visualized by a small cross. A recorded video of the performance[13] was then shown to a group of 31 people, none of whom was familiar with the EyeHarp. All participants reported at least a basic level in playing a musical instrument. Before showing the video performance, the audience was informed that the EyeHarp is a Gaze-Controlled digital musical instrument that consists of two different layers allowing the user to construct chords and arpeggios, control the harmony and play melodies. After having watched the video, the participants responded to a questionnaire. The questionnaire included questions intended to identify the profile of the listener (age, sex, music education, familiarity with DMIs and eye tracking technology) and questions exploring the evaluation criteria proposed by Barbosa and Calegario (2012). All responses were given in the form of linear scale from 1 to 5. Thirty-one people (6 women) of average age 30.5 years (standard deviation 5.8) responded to the questionnaire. Procedures were positively evaluated by the Parc de Salut MAR - Clinical Research Ethics Committee, Barcelona, Spain, under the reference number: 2013/5459/I. Participants responded questions for 6 evaluation criteria:

- Cause comprehension: were the available input gestures clear? (1: not at all. 5: very clear)

---

[13]Available online at https://youtu.be/dS5QkIgK0NY

- Effect comprehension: were the available control parameters clear? (1: not at all. 5: very clear)

- Mapping comprehension: was the connection between the input gestures and the control parameters clear? (1: not at all. 5: very clear)

- Intention comprehension: how well did the system allow the user to express his musical intentions? (1: not at all. 5: very well)

- Error comprehension: if there had been errors in the performance, would they have been noticeable? (1: not at all. 5: very noticeable)

- Enjoyment: how much did you enjoy the performance? (1: not at all. 5: a lot)

3.5.2   Performer Perspective

Quantitative Evaluation

The performer perspective evaluation was carried out with written informed consent from eight participants in accordance with the Declaration of Helsinki. Procedures were positively evaluated by the Parc de Salut MAR - Clinical Research Ethics Committee, Barcelona, Spain, under the reference number: 2013/5459/I. Participants (7 male, 1 female) with mean age of 34 years (SD 6.7) participated in a single-session quantitative evaluation task. All participants had some musical instrument playing experience. The quantitative evaluation consisted of a set of tasks using both the step sequencer and melody layer. Apart from one subject, no participant had previous experience with the EyeHarp DMI.

The Eyetribe low-cost commercial eye-tracker was used for acquiring the raw gaze data. Participants were comfortably seated at approximately 60 cm away from a 15.6 inches laptop screen placed at eyes level. All participants calibrated with 9 calibration points and 800 ms of sample and transition time. All participants achieved a 5-star calibration quality in the Eyetribe calibration software (expected visual angle accuracy = 0.5°) . A set of M-Audio AV40 self-amplified speakers were connected to the laptop audio output. The ASIO4ALL low latency driver was used, providing an audio output latency of 7 ms. The EyeHarp application was sending MIDI messages through loopBe1 virtual MIDI port to Reaper Digital Audio Workstation (DAW)[14], running a piano sound module for the Step

---

[14]http://www.reaper.fm/, last accessed on 16/3/2015

Sequencer layer and a clarinet sound module for the Melody layer. Gaze data were recorded in the EyeHarp application, whereas MIDI data were recorded in the Reaper DAW.

Step Sequencer layer evaluation   The step sequencer layer evaluation task consisted of constructing arpeggios with varying number of buttons in the step sequencer grid. All arpeggios were constructed three times. The first time the gaze pointer was hidden and no magnification method was applied (basic method). The second time the gaze pointer appeared along with additional focus point (gaze feedback method). The third time the gaze pointer was hidden and the described magnification method was applied (magnification method). In all cases when the gaze was detected inside a button, the fixation point was turning green. Figure 6 shows the 3 different arpeggios the participants were asked to construct in each of the 3 tasks. In the first task the grid size was 8x8, in the second 12x12 and in the third 16x16. In all cases, the participants were asked to correct all possible mistakes. The time to complete each task was measured.

Melody layer evaluation   Four different tasks of increasing difficulty were designed. Users practiced for about 2 minutes before recording 3 repetitions of each task. At the beginning of each task an arpeggio was constructed in the step sequencer layer that served as a metronome. Figure 7 shows the melodies the participants were asked to perform for each task: a scale in both directions, a scale with repeated notes, "twinkle twinkle little star", and a music exercise with a melody and a chord progression.

Qualitative Evaluation

After the quantitative evaluation session participants filled in a questionnaire. Participants responded (in a linear scale from 1 to 5) to the following questions:

- How much previous practice and training does the performer need for performing with the instrument, when compared to a traditional musical instrument? (1: no practice required. 5: extensive practice required)

- How much control does the performer have on the musical output? (1: restricted (equivalent to a DJ). 5: extensive musical control that allows expressive performance.

- How much real-time feedback (e.g. visual, auditory) does the user receive from the system? (1: low feedback. 5: high, multimodal feedback)

- How tiring is it to play music with your eyes when compared to the hands? (1: not tiring at all. 5: very tiring)

- Is it hard to play in tempo with your eyes when compared to hands? (1: equally hard. 5: much harder.)

- Which approach between the magnification lens and the fixation points do you consider more user-friendly? (1: I prefer the fixation points. 5: I prefer the magnification lens)

All questions were verbally explained to the participants. If anything seemed unclear to the participants they were free to ask for questions, which were clarified orally. In the first question, it was orally clarified that by the phrase "performing with the instrument" it is meant to achieve some basic, but rewarding interaction with the instrument. By the response "1: no practice required" we refer to the practice required to achieve a rewarding performance in a gaming music interface, like the guitar hero of Microsoft Xbox. By the response "5: extensive practice", we refer to the practice required to achieve a rewarding performance in a musical instrument that is considered to be difficult to learn, like the violin. Similarly, regarding the second question, it was clarified that by the response "5: extensive musical control that allows expressive performance" we refer to the control offered by an instrument like the violin. In question 4, it was orally clarified that users should respond "1: not tiring at all" if they consider it equally tiring as playing with the hands.

3.5.3 Results

Audience Perspective

Figure 8 shows the average responses and the corresponding standard deviation across all participants. The responses of the audience can be summarised as follows: The available input gestures were clear (average = 3.9, $\sigma = 0.87$). The available control parameters were clear (average = 3.8, $\sigma = 1.04$). The connection between them was clear (average = 3.7, $\sigma = 1.34$). The system allowed the user express his musical intention very well (average = 4.2, $\sigma = 0.76$). Errors in the performance would have been noticeable

(average = 3.1, σ = 1.06). Finally the audience enjoyed the performance a lot (average = 4.3, σ = 0.84).

Performer Perspective

Step Sequencer Layer   Figure 9 shows the average number of selections per minute across the 7 participants with no previous experience with the interface, for each task. The results obtained by the experienced user (M32) are shown separately at the same graph. The average number of selections per minute value is computed by dividing the number of required selections in each task by the time to complete the task.

In all tasks the experienced user performed about 2 to 3 times faster than the average speed across the native users. The best average performance (selections per minute) in the case of the 12x12 and 16x16 grid was achieved with the gaze feedback method. In the case of the 8x8 grid task, it was achieved with the basic feedback method. The lowest standard deviation value was achieved in all tasks with the magnification method.

Melody Layer   Figure 10 shows for each task and participant the percentage values of the notes played according to temporal accuracy. These values sum 100%, as they correspond to the temporal accuracy of played notes along with the omitted notes. In dark brown appears the percentage of accidentally played notes and in light brown appears the number of pauses made in each task. As pauses we refer to the cases where the participants stopped for one or more bars, in order to continue playing in tempo. The percentages are calculated by dividing the number of each value with the total number of selections that should be made in the task. The last column of each task corresponds to the average value across all participants, excluding the experienced participant. In figure 10, the code number of each participants was given by considering their sex, age and level of playing music in a scale from 1 to 5 (1: not playing any instrument, 5: professional level). For example user M48_4 is a 48 year old man with semi-professional level in playing music.

In all tasks the experienced user played around 20% more notes in tempo than the novice users, performed less accidental notes and no pauses.

Qualitative Evaluation   Figure 11 shows the average and standard deviation of the responses of the participants in the performer's evaluation.

3.5.4   Discussion

In the present study an evaluation of the proposed digital musical instrument has been conducted. This evaluation has been conducted both from the audience and the performer perspective. According to the audience's evaluation responses, the EyeHarp digital music instrument offers a transparent correspondence between input gestures and the produced sound, i.e. participants in the study average rating of their understanding of the cause (cause comprehension), the effect (effect comprehension), and gesture-sound correspondence (mapping comprehension) was greater than 3.5 out of 5 (see Figure 8). This denotes a high level of transparency and comprehensibility in the actions and their relationship with the produced sound of the proposed music instrument. According to Hunt et al. (2002) and Arfib et al. (2005), these properties (transparency and comprehensibility) are positively correlated with the capacity of an instrument to allow the musician to produce expressive performances, and to engage the audience in the performances. Nevertheless, the obtained standard deviation for the gesture-sound correspondence (mapping comprehension) evaluation (SD=1.33) indicates that some participants did not fully understand this correspondence. The standard deviation was smaller for the case of the cause and effect comprehension. Even though the EyeHarp being a diatonic DMI in which dissonant notes are very seldom produced, average audience evaluation of the error comprehension was high (i.e. 3.1). This again indicates a good understanding of the performer actions and corresponding produced music. All audience participants declared that they enjoyed the performance (average 4.3 out of 5). Most participants agreed that the interface allowed the performer to express his musical intentions (average 4.2 out of 5.0) which may be interpreted as an indication that the EyeHarp can allow the user to produce expressive performances.

Regarding the results of the evaluation from the performer's perspective, in the first task of the qualitative evaluation of the Step Sequencer Layer (i.e. the 8x8 grid task) it was achieved the best average time per selection. The resulting average time for the case of the 12x12 grid was almost double of the average time for the 8x8 grid task. This was expected, as small targets are harder to select. However, in the case of the 16x16 grid task the average selection time was less than the the average for the 12x12 grid task. This can be explained by the fact that most of the notes in the 16x16 grid task were adjacent notes, which makes the visual search task easier.

The 8x8 grid arpeggio task can be compared to typical dwell-time eye-

typing task, where the notes are replaced by characters. As seen in figure 9, the average number of notes per minute in the 8x8 grid is close to the average number of characters per minute in dwell-time typing systems (17 chars/min according to Hansen et al. (2003)).

In the case of the 8x8 grid the gaze feedback method produced the same results as the basic method, where the only visual feedback to the user is the brightening of the focus point at the center of the attended button. This result may be explained by considering the size of the 8x8 buttons: given their big size there was no difference with the two methods. On the contrary, in the case of the 12x12 and 16x16 grid, when the detected gaze coordinates were given as visual feedback, along with additional focus points, the performance (number of selected buttons per minute) increased with the gaze feedback method.

The experienced user participating in the study completed all the tasks of the step sequencer on average 2.8 times faster than the rest of the users. The difference is even higher in the case of the gaze feedback method. As concluded by Majaranta and Bulling (2014), if the user tries to look at the detected gaze coordinates, he may end up chasing the detected point, as it always is a few pixels away from the point he/she is looking at. It requires practice to learn how to take advantage of the visual feedback provided by the cursor in order to compensate for small calibration errors by adjusting the gaze point accordingly to bring the cursor onto an object. The experienced user clearly took more advantage of the gaze feedback than the non-experienced users. The difference in performance between the experienced user and the non-experienced ones may show that the EyeHarp is, similarly to traditional music instruments, an instrument in which practice play an important role.

The magnification method always performed worse than the gaze feedback method and only in the case of the 12x12 grid the obtained results were better than those obtained by the basic selection method. However, the magnification method always showed the lowest standard deviation on the number of selections per minute. This might explain why, as shown in figure 11, the users show a preference for the magnification method over the gaze feedback method. The gaze feedback method might not be appropriate for novice users.

All in all, the evaluation of the step sequencer layer, confirmed all results reported by similar gaze controlled systems in which selecting targets using dwell-time selection method is required (Hansen et al. (2003); Majaranta

and Rih (2007)): (i) There is a steep learning curve in gaze interaction, (ii) magnification methods help in selecting small targets, and (iii) gaze visual feedback improves the performance of experienced users.

Figure 10 clearly shows that in the melody layer the experienced user (M32) achieves better temporal accuracy than any other user. This is an indication that there is a learning process involved for adapting to use the melody layer. Nevertheless the number of accidental notes produced by the experienced user in tasks 2 and 3 are close to the average values across novice users. This indicates that the accidental notes are mainly caused by poor tracking accuracy, and not by the skill of the performer.

The number of omitted notes is higher in the tasks that require playing consecutively the same note (tasks 2 and 3). This is due to the behavior of the button responsible for note repetition: if a fixation is performed in the inner area of the pie but outside the "repeat note" button, the button disappears. In addition, due to noisy gaze tracking data, the user may be focusing on the center of the repeat button but the initial detected gaze point may fall outside the button area.

Although the tasks were designed with increasing difficulty, the average performance in the first task was similar to the average performance in the last task. This may be due to the training effect which compensates the different difficulty levels of the tasks. The last task is the most demanding, as it requires changing the chords along with the melody. A high number of accidental notes were observed during this task (as shown in Figure 10). This is due to he fact that the the positions of the chords and the notes are placed diametrically opposite in the interface.

The participants in the performer's perspective evaluation responded that the practice required to play the EyeHarp is comparable to the practice required to play a traditional musical instrument of average difficulty (3 out of 5 on average). The same response was given on average on the question about the the control the user has over the musical output (average value 3.1 out of 5), meaning that the control over the output is equivalent to that of a musical instrument that offers average control over the musical output.

The real-time feedback was rated high by most performers (average 3.9 out of 5). Most performers agree that playing music with the eyes is more tiring than playing with the hands (average 3.6 out of 5). Playing in tempo with the eyes is considered to be harder than playing with the hands (3.2 out of 5). Summarising the above responses, we could conclude that performing music with the eyes is more difficult that performing with traditional

means. Nevertheless, learning the EyeHarp gaze-controlled musical instrument wouldn't be harder than learning a traditional musical instrument.

The performer perspective evaluation was conducted with people with experience in playing musical instruments and no disabilities. In order to evaluate the EyeHarp in a more realistic setting, we would have required to test it with locked-in syndrome patients. This, we believe, should be done in the future, and we have started looking for possible participants.

As summary, we have presented and evaluated the EyaHarp, a new gaze-controlled digital musical instrument. The system was evaluated from the performer and audience perspective. The obtained results indicate that, similarly to traditional music instruments, the proposed digital musical instrument allows to produce expressive performances both from the performer and audience perspective. The participants in the evaluation from the perspective of the performer responded that the practice required to master the EyeHarp DMI is similar to the average practice required to master a traditional musical instrument of average difficulty. The steep learning curve of the instrument is also reflected on the quantitative data, when comparing the performances of the experienced user with the novice users.

The cost of eye-tracking technology decreases every year. The last 5 years the cost of commercial eye-trackers has been reduced more than 10 times. Eye-tracking is slowly being incorporated in common place laptops, tablets and mobile phones. Such devices would allow many users, including users with motor disabilities, to have access to gaze-controlled applications, including the EyeHarp DMI.

The pEYE interface in the melody layer, provides a solution to the Mida's touch problem making it possible to play melodies in-tempo when the gaze of the user is used as the only input. If the physical abilities of the user allow it, other selection techniques like blinking, using physical buttons or blowing could be considered. If such selection methods were utilized, the user would be able to freely visually search the screen without triggering any undesired notes. This would allow increasing the number of available notes on the screen, as the central (neutral) area of the melody layer wouldn't be necessary. As future work, it would be interesting to compare the performance -in terms of overall usability, temporal accuracy and speed- of such an interface with the current version of the EyeHarp. The advantage of the screen button selection method may be that just one action is required to play a note: looking at the selection area. This might allow playing faster

than in the case of using an independent clicking method which requires two actions (i.e. looking at the selection area and clicking). On the other hand, using an independent clicking method might allow placing more notes on the screen and might allow better temporal accuracy.

Probably the main target group of the proposed DMI is that of people diagnosed with Amyotrophic Lateral Sclerosis (ALS). ALS is a progressive neurodegenerative disease that affects nerve cells in the brain and the spinal cord. Individuals affected by the disorder may ultimately lose the ability to initiate and control all voluntary movements. Nevertheless, muscles responsible for eye movement are usually spared until the final stages of the disorder (LAWYER (1953); Kiernan et al. (2011)). A large number of studies have shown that music playing provides a variety of benefits (e.g. cognitive, psychological) (e.g. Hays and Minichiello (2005)). The EyeHarp DMI gives the opportunity to ALS patients to have access to such benefits. This could have a big positive impact in the quality of life of ALS patients -musicians or not-, by providing them the possibility of playing a musical instrument.

3.6   Conclusions

In this section we have presented the EyeHarp, a new gaze-controlled digital musical instrument. Initially we evaluated the temporal accuracy when playing melodies using a simple dispersion-based algorithm. The finding propose that the asynchrony is correlated with the distance between consecutive fixations.

Driven by this indication we proposed a new fixation detection algorithm and we compared it in a study involving 7 healthy subjects. The new algorithm (dispersion-based of varying time window) was compared against a dispersion-based and a velocity based algorithm. The results indicate that the new proposed algorithm provides better overall temporal accuracy than commonly used fixation detection-algorithms.

Finally we evaluated EyeHarp DMI from the audience's and the performer's perspective. The responses of the audience indicate a high level of transparency and comprehensibility in the actions and their relationship with the produced sound of the proposed music instrument. According to Hunt et al. (2002); Arfib et al. (2005) ,these properties are positively correlated with the with the capacity of an instrument to allow expressive performances. Regarding the quantitative evaluation performed from performer's perspec-

tive, the experienced user performed much better in all tasks. He performed even better when the gaze feedback selection method was active, in which the user is aware of the detected gaze point. In average, participants in the evaluation from the performer's perspective think that the practice required to play the EyeHarp is comparable to the practice required to play a traditional musical instrument of average difficulty. The same response was given in the question about the control the user has over the musical output. In summary we could conclude that performing with the EyeHarp is more difficult than performing with traditional means. Nevertheless, learning the EyeHarp wouldn't be harder than learning a traditional musical instrument. The steep learning curve of the instrument is also reflected on the quantitative data, when comparing the performances of the experienced user with the novice users.

Brain-Controlled Music Interfaces

4.1 Introduction

An interface in which the brain activity is used as input, is called a Brain-Computer Interface (BCI). In this chapter we will explore ways that might enhance musical expression using the brain activity as measured by electroencephalography (EEG). Electroencephalography is the most common technique for measuring brain activity in order to build a BCI for its low cost and high temporal resolution of a few milliseconds. Other alternatives for measuring brain activity include near-infrared systems (fNIR) (e.g. Coyle et al. (2007)), magnetoencephalography (MEG) (e.g. Mellinger et al. (2007) and functional magnetic resonance imaging (fMRI) (e.g. Sitaram et al. (2007). fMRI and MEG require expensive, heavy and unportable equipment. fNIR on the other hand is cheaper and more compact. As both fNIR and fMRI measure the cerebral blood flow, they have poor temporal resolution. Extensive BCIs reviews can be found in Wolpaw et al. (2002), McFarland and Wolpaw (2011), Edlinger et al. (2012) and Kaur et al. (2012).

The EEG signal consists of a number of EEG channel. Each channel represents the difference in voltage between two electrodes placed on the scalp of the subject. Normally the electrodes are places in standard positions defined by the international 10-20 system (Niedermeyer and da Silva (2005)). The main EEG device used in this PhD was the Emotiv Epoc headset. Figure 4.1 shows the labels assigned to each electrode position, with the

71
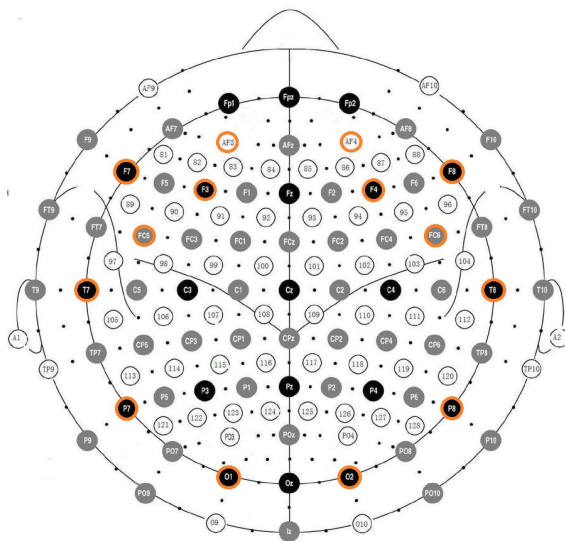
Figure 4.1: The international 1020 system. The locations provided by the Emotiv Epoc are highlighted.

positions provided by Emotiv Epoc highlighted.

Zander et al. (2010) proposed three distinct BCI categories, depending on the way the system derives its output. (i) An active BCI derives its outputs from brain activity which is directly consciously controlled by the user, independently from external events, for controlling an application. (ii) A reactive BCI derives its outputs from brain activity arising in reaction to external stimulation, which is indirectly modulated by the user for controlling an application. (ii) A passive BCI derives its outputs from arbitrary brain activity without the purpose of voluntary control, for enriching a human-computer interaction with implicit information.

Another distinction we can make is based on the underlying neuromechanism:

Event Related Potentials An event-related potential (ERP) is the measured brain response that is the direct result of a specific sensory, cognitive, or motor event (Nunez and Srinivasan (2006)). It is one of the most robust ways to built a BCI. The most common example is the P300 speller. The user thinks of a specific letter he wants to write and looks at a matrix of

flashing letters. About 300ms after the desired letter flashes, a positive pick appears in the EEG signal. P300 evoked potentials appear with any kind of stimulus, including auditory events. Other evoked potentials include Visual Evoked Potentials or Error Related Negativity. Event related potential BCIs are reactive BCIs.

Feedback based: Slow Cortical Potentials (SCP), Sensorimotor Rhythm (SMR) BCIs is a skill that users can practice. The brain itself has a significant plasticity in acquiring new skills. When appropriate feedback is given, the user can be trained to consciously control the extracted EEG features. SCP and SMR manipulation are the most successful examples of feedback based BCIs.

Slow Cortical Potentials are potential shifts in the EEG signal of 2 Hz or less. It is not localized in the cortex. In general a negative shift is a sign of readiness/mobilization, while positive shift indicates an ongoing cognitive task or inhibition of neuronal activityHinterberger et al. (2004a). In every day life negativity could appear when they call us by name (readiness) and positivity when we try to understand what they are telling us. The Thought Translation device proposed by Birbaumer et al. (2000) is based on SCP manipulation and is probably the most robust and simple current active BCI. SMR (or mu rhythm) manipulation is achieved by measuring the activity over the sensorimotor cortex. Both SMR and SCP can be achieved with auditory or visual feedback (Nijboer et al. (2008b)).

Motor Imagery By measuring the activity of separate regions on the sensorimotor cortex, and after applying the appropriate spatial filters, extracting the appropriate features a classifier is trained to distinguish between a number of imagery movements. For example in a hybrid gaze-brain interface proposed by Lee et al. (2010) the user "grabs" the items he is focuses on.

More invasive techniques lead to more robust and accurate interfaces. In a study conducted by Hochberg et al. (2012), a woman was able to control a robotic arm in a broad space to grab a bottle and drink. To achieve this, a microelectrode array was attached on a small, local population of motor cortex (MI) neurons.

Response to mental tasks Users are instructed to perform a mental task, like solving a equation or performing music imageryBlokland (2009).

Then following a similar procedure described in motor imagery, a classifier is trained to distinguish between the given mental states.

4.1.1 State of the Art in Brain Computer Music Interfaces

The idea of producing sound using the electro-physiological measurements of the brain activity, ages back to 1934, when Adrian E.D. and Matthews B.H.C E.D. and B.H.C. (1934) sonified the signal coming from "electrodes applied to the head" while studying the *Berger rhythm* [1]. The most straight-forward way of producing sound or music through the EEG signal is by mapping features extracted from the EEG signal to sound or music param-eters. EEG sonification is a way of representing mental states using auditory output. The application of different sonifications vary (an extensive review was made by Väljamäe et al. (2013)). (i) Monitoring applications inform a third person about the state of the subject (e.g. during surgery). (ii) A similar offline scenario is offered by diagnostic applications (e.g. sonifica-tion of sleep states). (iii) Neurofeedback applications inform the user about his own mental state, normally aiming to achieve a specific state of mind (e.g. meditation or concentration). (iv) Brain-Computer Interface feedback and communication applications. As an example, Teitelbaum (1976) used the alpha waves (EEG signal band-passed to the 8-13 Hz range) and other biological signals to control an electronic synthesizer.

The first use of EEG for a musical composition was conducted by Alvin Lucier, in 1965 performing the musical piece *Music for Solo Performer*. The amplified EEG signal was driven to loudspeakers. Various percussions attached to the loudspeakers would resonate producing the first artistic EEG-sonification (Lucier (1976)). Most sonification approaches map the power of different frequency bands of the EEG into control parameters. Table 4.1.1 shows the mental association related to different power bands. An example of such a sonification was presented by Hinterberger and Baier (2005), in which each of 6 frequency bands of the EEG signal is assigned an instrument of a MIDI device. The EEG signal from 1 (Cz - basic setup) or 3 electrodes (Cz, C3 and C4) is filtered into various frequency bands result-ing in 17 control channels. These channels controlled either the pitch or the volume of 11 midi instruments, resulting in a "Orchestral Sonification". Ten

---

[1] "a rhythmic oscillation of potential at a frequency of 10 cycles per second . . . detected in the human subject by electrodes applied to the head . . . present when the subject lies quietly with eyes closed and disappearing when attention is fully occupied" (Hans Berger, 1929)

EEG Rhythm	Frequency Band	Mental Association
Delta	$\delta <= 4\text{Hz}$	Sleep
Theta	$4\text{Hz} < \theta \leq 8\text{Hz}$	deep meditation, deep relaxation, drowsiness
Alpha	$8\text{Hz} < \alpha \leq 13\text{Hz}$	Relaxed wakefulness, closed eyes
Beta	$13\text{Hz} < \beta \leq 30\text{Hz}$	Awake, alertness, mental activity
Gamma	$\gamma > 30$	Hyper-awareness, stress, anxiety

Table 4.1: Bands of EEG activity and associated mental states for a healthy young adult

subjects were instructed to try to focus on different sounds of the orchestra and the difference of the mean value (t-test) for each input parameter was computed. This difference was significant in the case of some subjects -especially in the alpha power region-, indicating that some subjects were able to control these parameters by focusing in different instruments, without any training.

Wu et al. (2009) attempted one more direct EEG to midi sonification where the pitch, duration and intensity of the piano notes are controlled by simple EEG features.

The notion of feedback might be crucial in a BCI. It has been proven than when appropriate feedback is given, subjects are able to control their Slow Cortical Potentials. Thilo Hinterberger et al. in Hinterberger et al. (2004b) examined whether SCP regulation can be achieved when auditory feedback is given. Audio feedback was compared with visual and combined (audio+visual) feedback. Fifty-four different subject were trained over 3 sessions to control their SCP, 18 in each category (visual, auditory, combined). Although the subjects were able to better control their SCP when visual feedback was given, the authors conclude that SCP self regulation is feasible with auditory feedback as well.

In Nijboer et al. (2008b) visual and auditory feedback were compared for controlling the sensorimotor rhythms. SMR desynchronization (non-motor imagery) produced a bongo sound, while SMR synchronization produced a harp sound. 16 participants were divided in two groups of auditory and visual feedback. After three training sessions both groups achieved accuracy more than 70% in choosing two different visual or auditory targets.

Another example of direct EEG sonification was performed by Sebastian Maella et al Mealla et al. (2011). Groups of two people tried to reproduce

a pre-recorded musical piece by interacting with the Reactable -a musical tabletop interface- and electrophysiological measurements of one of them. Heart beat was controlling the tempo on the Reactable interface (Jordà (2010)) while the alpha-theta bands (4-12Hz) of one dry electrode placed on the frontal lobe was driving the a direct sonification (in the audible sound frequency spectrum) of a sound puck on the Reactable. Both users were able to manipulate the pucks with their hands. In 5 out of 16 groups, the EEG data were pre-recorded (placebo group). The subjects in the non-placebo group reported more confidence and control during the interaction process supporting the feasibility of physiology-based interaction in multi-modal interfaces for collaborative music.

Although there is evidence that even with direct sonification approaches, when appropriate auditory biofeedback is given a user could be trained to control a auditory BCI, this might be particularly difficult to accomplish. For that reason many researchers are trying non-direct EEG sonification, based on various EEG potentials and trying higher-level mappings.

Ben Swift et. al in Swift et al. (2007) (2007) describes a *Mind Attention Interface* based on detecting the functional connectivity [2] that appears during musical processing. The amount of measured functional connectivity is a measurement of user's attention and affects navigation in a space of Harmonic Triads, generating a chord progression.

E. Miranda in Miranda et al. (2003) developed one more interface based on the user' s attention - the "brain soloist"-. This could be described as a music imagery driven interface. In this EEG BCI, the system is initially trained to distinguish between two mental states: passive and active listening (whether the user sings the tune mentally or not). The system consisted of an 128-channel Geodesic System. After applying appropriate Laplace Filtering, an autoregressive algorithm was utilized to extract the features that trained a neural network. The achieved accuracy was reported to be more than 95% within 3 subjects. The performer listens to a steady rhythmic part and riffs are played sporadically. After one riff is played the system checks if active listening is detected. In that case a variation of this riff is generated and played back. Otherwise the same riff is played unchanged until active listening is detected.

---

[2]Functional connectivity describes the task-dependent connections between distinct brain functional regions. It can be detected by observing synchronous activity in different regions of the brain.

Blokland (2009) explores the feasibility of a BCI based on music imaginary. In such a BCI the system first has to be trained to distinguish whether the user is internally listening to a certain melody. Using the Biosemi Active-2 64 electrodes system and a linear logistic regression classifier just a poor 62% maximum classification rate was achieved.

In 2005 Miranda et. al developed a system where EEG patterns control music generating algorithms Miranda and Boskamp (2005). The generative component of the system employs artificial intelligence techniques to imitate different musical styles. In particular the demonstrated system activates generative rules of two different musical styles depending on the whether the system has dominant low or high frequencies components. The tempo of the composition is controlled by a signal complexity analysis algorithm. The authors indicate the need of a biofeedback system that would train the performed to gain control over the biological signals. Later Miranda et al. (2008) proposed a generative music approach based on constraint satisfaction techniques. In such a system the rules of classical harmony can be expressed as mathematical relations between the notes that are expressed as integers. The system then searches for an acceptable solution that does not violate any constraint. As a Constraint Satisfaction Problem might have more than one solution, EEG data could drive the automated music generation, while the constraints guarantee a melodic outcome.

Makeig et al. (2011) proposes a musical BCI that aims to sonify the emotions of the performer. The system is trained to distinguish between 5 emotions that the performer was instructed to feel while listening to a musical parts and afterwards the performer regenerates these parts by bringing himself in the desired emotional state.

Examples of P300 based systems, based on visual stimuli, for composing music are those presented by Grierson (2008) and Hamadicharef et al. (2010). In such a system the name of the notes in a musical scale flash on a computer screen, and P300 evoked potentials provoke the intention of the user to select a note. In a similar way a P300 online step sequencer was presented by Chew and Caspary (2011). The BCI2000 software was used for presenting the flashing notes in the sequencer grid and a Stepwise Linear Discriminant Analysis that comes with the BCI2000 software was utilized to create the P300 classifier 4.2. Evaluation over 10 participants resulted to an average accuracy of 86%.
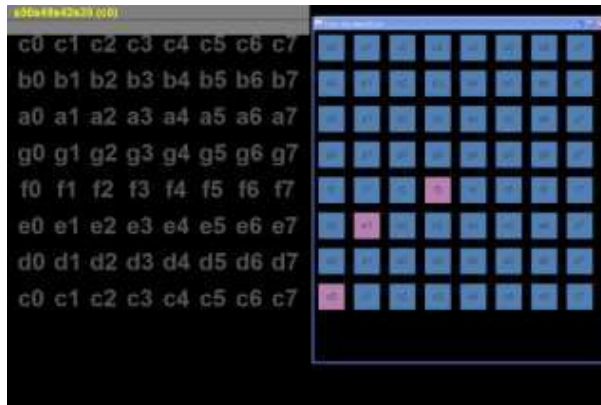
Figure 4.2: MusEEGk. A P300 based step Sequencer. The left matrix is used for selecting the notes while the right matrix displays the notes selected

4.2 Motor Imagery

Motor imagery is another mechanism used in BCIs. In this case the user imagines of moving a part of his body, such as a hand or a leg. This has been shown to produce increased activity in the region of the motor cortex at the frequency band 7.5 Hz to 12.5 Hz. This electrical activity is commonly referred as mu-rhythm. When provided with feedback, users can be trained to voluntarily control the power of mu-rhythm. The most typical application is moving a cursor in two dimensions on a computer screen. This idea was first presented by Wolpaw et al. (1991). Large mu rhythm amplitude moved the cursor up and low amplitude down. After several weeks of training, subjects learned to control mu-rhythm quickly and accurately (hit targets within 3 seconds). Alternatively, auditory feedback can produce similar results. In a study performed by Hinterberger et al. (2004a), the frequency and amplitude of the peak of the mu rhythm power were mapped to different parameters of a sonification through midi. In specific the amplitude of the mu rhythm was mapped to the amplitude of the produced sound, and the frequency of the detected mu rhythm peak was mapped to the pitch of the sound. Five out of ten subjects achieved in a single session a correct response rate of 60%.

McCreadie et al. (2013) conducted an experiment on 20 participants comparing visual with different types of auditory feedback in mu-rhythm manipulation training. The results indicate that stereophonic auditory feedback

with broadband noise provides the most robust auditory feedback method. Although with visual feedback the participants achieved more accurate manipulation of their mu-rhythm, with auditory feedback a positive learning curve was observed.

The study presented on this section was presented at the Proceedings of the Fifth International Brain-Computer Interface Meeting 2013 (Vamvakousis and Ramirez (2013)).

4.2.1   Playing Melodies using mu-rhythm

In a study involving 6 male right-handed subject, the classification accuracy of real and imagery toes movement over resting state was studied. This system was then used to move the pitch of a virtual instrument one step up or down in the c major scale. Results were published in Vamvakousis and Ramirez (2013).

Continuous feet movement has been reported to cause Event-Related Synchronization (ERS) around the FC3 and FC4 standard positions of the EEG in the high alpha and low beta band (Pfurtscheller et al. (2006); Jeon et al. (2011); Wang et al. (2009)).

Materials and Methods   Six male right-handed healthy subjects, of average age 34 years, took part in one real and one imagery movement experiment. Each session consisted of the following steps: Twelve trials (6 movement and 6 non-movement) of 10 seconds each were randomized. When an arrow pointing upwards appeared on the screen continuous real (in the case of the real movement session) or imagery (in the case of imagery movement session) toes movement should be performed for 10 seconds, while if the arrow pointed downwards they should stay relaxed for 10 seconds. The subjects were instructed to avoid any unnecessary muscular activity.

The Emotiv Epoc EEG headset was used to capture EEG data. The headset was placed in a way that the frontal electrodes captured the motor cortex region. Using OpenVibe software, in an on-line scenario, the signal was filtered in the 10-17Hz band using a fourth order Butterworth band pass filter. The data recorder for each subject were used to train an LDA and a third degree polynomial SVM classifier. A moving 2 seconds window of hop size 100 ms was applied on each channel and the power of each window was used as a feature for the classifier. During the recording the sum of the power of all channels were added and plotted on the screen. During real or
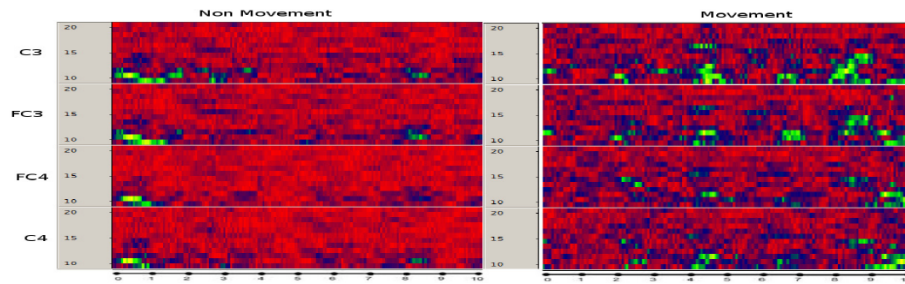
Figure 4.3: ERS observed in the motor cortex of a subject during real feet movement

imagery toes movement the objective of the participants was to maximize the displayed power value. The classifiers were trained using different time intervals within the 10 seconds period of each trial.

The classifiers were trained using different time intervals within the 10 seconds period of each trial.

Results and Discussion   Toes movement was observed to cause a gradual increase of the computed power as opposed to the resting state, as a result of ERS in the sensorimotor cortex (see figure 4.3). In 4.2 the average 10-cross validation performance and variance for 6 subjects, for different time intervals is displayed. Looking at the table we can make the following observations: (i) The polynomial SVM classifier always outperforms the LDA classifier. (ii) Mu rhythm synchronization needs some time to develop both in the case of real and imagery movement. When the last 3 out of 10 seconds are used for the classification, the average SVM 10-fold cross validation performance is 91.65% in the case of real movement and 85.57% in the case of imagery movement. The overall performance falls when earlier intervals are used. This indicated that ERS takes some time to develop. (iii) Real toes movement resulted in stronger ERS than imagery movement. Although in the case of 7-10 s window with SVM polynomial classifier the 85.57% average performance indicates that an imagery movement based interface is feasible.

As a case study a simple musical application was designed, where the last 3 seconds of a 10 seconds movement or non-movement trial were used to control the contour of a melody. Initially the threshold of an LDA classifier is computed by asking the user to perform three 10-second long real movement and non-movement trials. Then every 10s the user performs a movement

Table 4.2: 10-fold cross validation performance average and variance for 6 subjects for different time intervals with an LDA and a

	Real Movement				Imagery Movement			
	Performance (%)		Variance (%)		Performance (%)		Variance (%)	
Time Interval	LDA	SVM	LDA	SVM	LDA	SVM	LDA	SVM
7-10	87.79	91.65	0.43	0.36	79.65	85.57	0.81	0.29
5-10	77.95	82.25	0.51	0.84	70.47	75.62	0.41	0.29
2-5	71.17	75.30	1.55	1.31	69.15	74.97	0.82	0.64
0-10	69.31	71.51	1.04	0.8	65.88	69.55	0.12	0.17

or non-movement trial depending on his intention. When movement is detected (value higher than the threshold) the melody moves up while in the opposite case, it moves down. Preliminary results on one subject indicate that the contour of the melody is controlled with enough accuracy.

4.3 An emotion estimation method through EEG signal

Introduction

Though music performances, the performer is expressing emotions. A review on research on communication of emotion in music performance was made by Juslin (2001). In BCMIs, where the brain activity is recorded, it is possible to detect emotions (e.g. Chanel et al. (2006); Choppin (2000); Horlings et al. (2008); Musha et al. (1997); Vourkas et al. (2000). There have been several approaches to EEG-based emotion detection, but there is still little consensus about definite conclusions. These emotions can be used to enhance the expressiveness of a BCMI, or of any music interface.

Alpha and beta wave activity may be used in different ways for detecting emotional (arousal and valence) states of mind in humans (more details later). Choppin (2000) propose to use EEG signals for classifying six emotions using neural networks. Choppin's approach is based on emotional valence and arousal by characterizing valence, arousal and dominance from EEG signals. He characterize positive emotions by a high frontal coherence in alpha, and high right parietal beta power. Higher arousal (excitation) is characterized by a higher beta power and coherence in the parietal lobe, plus lower alpha activity, while dominance (strength) of an emotion is characterized as an increase in the beta / alpha activity ratio in the frontal lobe, plus an increase in beta activity at the parietal lobe.

Bos (2006) describes an approach to recognize emotion from EEG signals measured with the Brainquiry EEG PET device. He uses a limited number of electrodes and trains a linear classifier based on Fishers discriminant analysis. He considers audio, visual and audiovisual stimuli and trains classifies for positive/negative, aroused/calm and audio/visual/audiovisual.

Takahashi (2004) uses a headband of three dry electrodes to classify five emotions (joy, anger, sadness, fear, and relaxation) based on multiple bio-potential signals (EEG, pulse, and skin conductance). He trains classifiers using support vector machines and reports the resulting classifying accuracy both using the whole set of bio-potential signals, and solely based on EEG signals.

Lin et al. (2010) apply machine-learning techniques to categorize EEG signals according to subject self-reported emotional states during music listening. They propose a framework for systematically seeking emotion-specific EEG features and exploring the accuracy of the classifiers. In particular, they apply support vector machines to classify four emotional states: joy, anger, sadness, and pleasure.

In this study, we describe an approach to detecting emotion from electroencephalogram signals measured with a (low-cost) Emotiv EPOC headset. We present to subjects auditory stimuli from a library of emotion-annotated sounds and record their response EEG activity. We then filter and process the signal in order to extract emotion-related features and apply machine learning techniques to classify emotional states into high/low arousal and positive/negative valence (e.g. happiness is a state with high arousal and positive valence, whereas sadness is a state with low arousal and negative valence). Our approach differs from previous works in that we do not rely in subject self-reported emotional states during stimuli presentation. Instead, we use a library of emotion-annotated sounds publicly available for emotional research. Figure 4.4 illustrates the different steps of our approach.

The study described in this section has been published in LNCS (Ramirez and Vamvakousis (2012)).

4.3.1   Methods

Data Collection

Subjects were instructed to look at a cross in the a computer screen and to remain seated during the experiment. Subjects listened to selected sounds from the IADS library of emotion-annotated sounds which is available for

Figure 4.4: Schematic view of the system.

emotion research (Lang et al. (1999)). Based on the annotations provided by the stimuli databases, we selected 12 sound stimuli situated in the extremes on the arousal-valence emotion plane: three positive/aroused, three positive/calm, three negative/calm, and three negative/aroused. The stimuli were selected to be as much as possible on the extremes of the two-dimensional emotion plane and as unanimous as possible, since we do not consider self-reporting information to cater for person-dependent deviations.

Initially, the subjects are informed about the experiment procedure and instructed to follow the usual guidelines during stimuli presentation (e.g. do not blink or move). Once this was done, 12 sound stimuli are randomly presented each one for five seconds and a 10 second silent rest is inserted between stimuli. The purpose of the 10 second silent rests is to set a neutral emotional state of mind in between stimuli.

Feature Extraction

In EEG signals the alpha (8-12Hz) and beta (12-30Hz) bands are particular bands of interest in emotion research for both valence and arousal (Niemic et al. (2002)). The presence of EOG artifacts (eye movement/blinking) is

most dominant below 4Hz, ECG (heart) artifacts around 1.2Hz, and EMG (muscle) artifacts above 30Hz. Non physiological artifacts caused by power lines are normally present above 50Hz Fatourechi et al. (2007); Coburn and Moreno (1988).

Thus, fortunately a byproduct of extracting the alpha and beta frequencies is that much of the noise present in EEG signals is considerably reduced. We apply bandpass filtering for extracting alpha and beta frequency bands. Using Fourier frequency analysis, the original signal is split up in frequencies in order to remove specific frequencies, before transforming back the signal with only the frequencies of interest. For this research, we apply the bandpass filter implementation provided by the OpenVibe software (Renard et al. (2010)).

From the EEG signal of a person, we determine the level of arousal, i.e. how relaxed or excited the person is, by computing the ratio of the beta and alpha brainwaves as recorded by the EEG. We measure the EEG signal in four locations (i.e. electrodes) in the prefrontal cortex: AF3, AF4, F3 and F4 (see figure 4.1). As mentioned before, beta waves are associated with an alert or excited state of mind, whereas alpha waves are more dominant in a relaxed state. Alpha activity has also been associated to brain inactivation. Thus, the beta/alpha ratio is a reasonable indicator of the arousal state of a person.

In order to determine the valence level, i.e. negative or positive state of mind, we compare the activation levels of the two cortical hemispheres. This is motivated by psychophysiological research which has shown the importance of the difference in activation between the cortical hemispheres. Left frontal inactivation is an indicator of a withdrawal response, which is often linked to a negative emotion. On the other hand, right frontal inactivation may be associated to an approach response, or positive emotion.

As mentioned before, high alpha activity is an indication of low brain activity, and vice versa. Thus, an increase in alpha activity together with a decrease in beta waves may be associated with cortical inactivation Niemic et al. (2002). F3 and F4 are the most used positions for looking at this alpha activity, as they are located in the prefrontal lobe which plays a crucial role in emotion regulation and conscious experience.

Although previous research suggests that hemispherical differences are not an indication of affective valence (feeling a positive or negative emotion), it has been suggested that it is an indication of motivational direction (approach or withdrawal behavior to the stimulus) Harmon-Jones (2003).
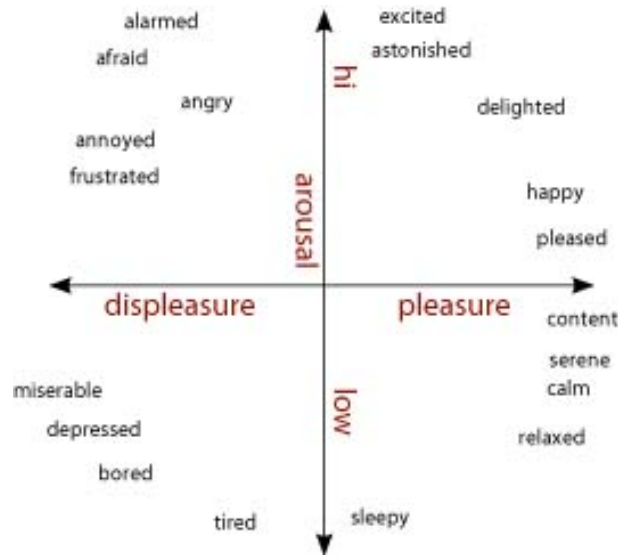
Figure 4.5: The Arousal-Valence plane, describing emotional states. Arousal is in the x axis and valence in the y axis.

In general, however, affective valence is related to motivational direction. Therefore, comparing hemispherical activation seems to be a reasonable method to detect valence. Thus, we estimate the valence value in a person by computing and comparing the alpha power a and beta power b in channels F3 and F4. Specifically,

$$valence = a_{F4}/b_{F4} - a_{F3}/b_{F3} \qquad (4.1)$$

Valence and Arousal Classifiers

In this section we describe our approach to training and evaluating classifiers for the task of detecting the emotional state of mind of a person given the person's observed EEG data. We approach this problem as a two 2-class classification problem. In particular, we apply machine learning techniques to classify high/low arousal and positive/negative valence emotional states. The obtained classifiers can be used to classify emotions such as happiness, anger, sadness, and calm. Figure 4.5 shows these emotions in the arousal/valence plane.

We are interested in inducing two classifiers of the following forms:

$$ArousalClassifier(EEGdata([t, t + c])) \rightarrow \{high, low\}$$

and

$$ValenceClassifier(EEGdata([t, t + c])) \rightarrow \{positive, negative\}$$

where $EEGdata([t, t+c])$ is the EEG data observed at time interval $[t, t+c]$ and $\{high, low\}$ and $\{positive, negative\}$ are the sets of emotional states to be discriminated. The results reported in this paper are obtained with c=1s and with increments of t of 0.0625s. For each subject in the EEG data sets we train a separate classifier.

In this paper we evaluate two classifiers, Linear Discriminant Analysis (LDA) Scholkopft and Mullert (1999) and Support Vector Machines (SVM) Cristianini and Shawe-Taylor (2000), for classifying an emotion state for each EEG segment. Linear discriminant analysis and the related Fisher's linear discriminant are methods used in statistics, pattern recognition and machine learning to find a linear combination of features which characterizes or separates two or more classes of objects or events. The resulting combination may be used as a linear classifier. LDA is closely related to regression analysis, which also attempt to express one dependent variable as a linear combination of other features. In regression analysis however, the dependent variable is a numerical quantity, while for LDA it is a categorical variable (i.e. the class label).

On the other hand, SVM is one of the most popular supervised learning algorithms for solving classification problems. The basic idea in SVM is to project input data onto a higher dimensional feature space via a kernel transfer function, which is easier to be separated than that in the original feature space. Depending on input data, the iterative learning process of SVM would eventually converge into optimal hyperplanes with maximal margins between each class. These hyperplanes would be the decision boundaries for distinguishing different data clusters. Here, we use linear and radial basis function (RBF) kernel to map data onto a higher dimension space. The results reported are obtained using the LDA and SVM implementations in the OpenVibe software Renard et al. (2010).

We evaluated each induced classifier by performing the standard 10-fold cross validation in which 10% of the training set is held out in turn as test data while the remaining 90% is used as training data. When performing the 10-fold cross validation, we leave out the same number of examples per
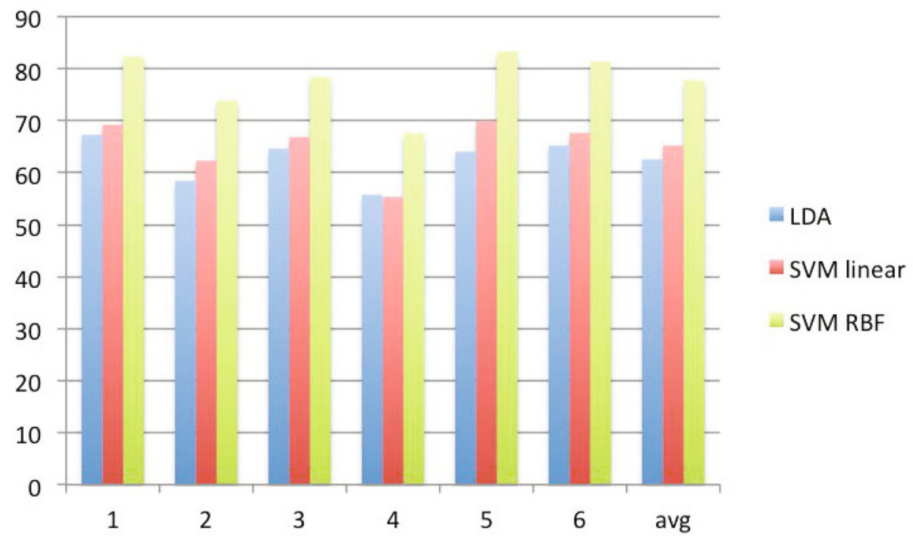
Figure 4.6: Classifiers (LDA, SVM with linear kernel, and SVM with radial basis function kernel) accuracies for high-versus-low arousal for all subjects.

class. In the data sets, the number of examples is the same for each class considered, thus by leaving out the same number of examples per class we maintain a balanced training set.

4.3.2 Results

Given that we are dealing with 2-class classification tasks and that the number of instances in each class is the same, the expected classification accuracy of the default classifier (one which chooses the most common class) is 50% (measured in correctly classified instances percentage). For the high-versus-low arousal, and the positive-versus-negative valence classifiers the average accuracies obtained for SVM with radial basis function kernel classifier were 77.82%, and 80.11%, respectively. For these classifiers the best subject's accuracies were 83,35%, and 86.33%, respectively. The correctly classified instances percentage for each subject and each learning method is presented in Figures 4.6 and 4.7.

4.3.3 Discussion

The difference between the results obtained and the accuracy of a baseline classifier, i.e. a classifier guessing at random confirms that the EEG data
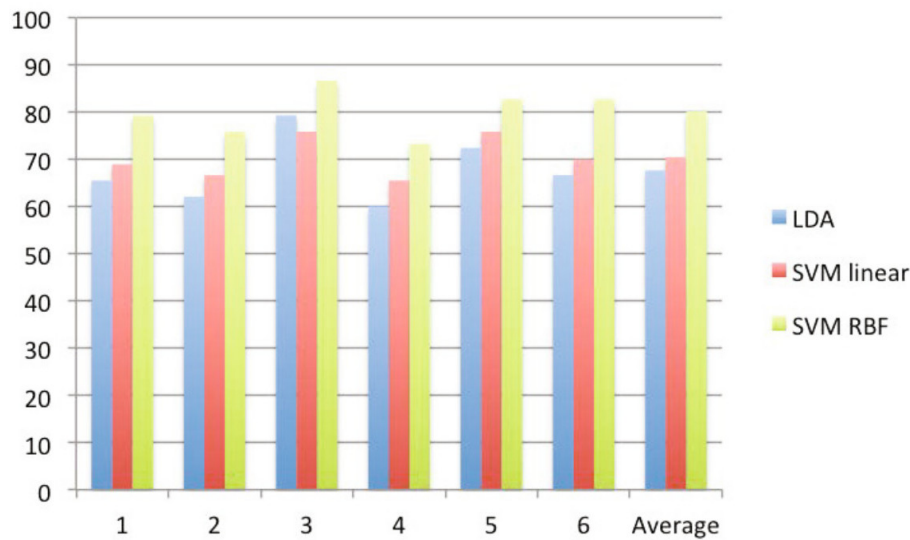
Figure 4.7: Classifiers (LDA, SVM with linear kernel, and SVM with radial basis function kernel) accuracies for positive-versus-negative valence for all subjects.

contains sufficient information to distinguish between high/low arousal and positive/negative valence states, and that machine learning methods are capable of learning the EGG patterns that distinguish these states. It is worth noting that both learning algorithm investigated (LDA and SVM) produced better than random classification accuracies. This supports our statement about the feasibility of training classifiers using the Emotiv Epoc for the tasks reported.

The accuracy of the classifiers for the same task for different subjects varies significantly, even using the same learning method. Subjects producing high accuracies with one learning method tend to produce high accuracies with the other learning methods. These uneven accuracies among subjects may be due to different degrees of emotional response between different individuals, or to the amount of noise for different subjects. In any case, it has been reported that there exists considerable variation in EEG responses among different subjects.

It is worth mentioning that in all the experiments performed we provided no self-assessment information about the emotional states by the subjects. This contrasts with other approaches (e.g. Lin et al. (2010)) where EEG data is categorized according to subject self-reported emotional states. Incorpo-

rating self-assessment information would very likely improve the accuracy of the classifiers.

We have explored and compared two machine learning techniques for the problem of classifying the emotional state of a person based on EEG data using the Emotiv Epoc headset. We considered two machine learning techniques: linear discriminant analysis and support vector machines. We presented the results of the induced classifiers which are able to discriminate between high-versus-low arousal and positive-versus-negative valence. Our results indicate that EEG data obtained with the Emotiv Epoc device contain sufficient information to distinguish these emotional states, and that machine learning techniques are capable of learning the patterns that distinguish these states. Furthermore, we proved that it is possible to train successful classifiers with no self-assessment of information about the emotional states by the subjects.

4.4 Combining Eye Tracking with Emotion Detection

In the previous section we have shown that *beta/alpha* frontal activity is correlated with emotional arousal, and the value computed by 4.1 is correlated with emotional valence. In this section we will describe how these these features were used in a concert setting, in which the estimated emotional state of the performer triggered predefined chord sequences. The Emotiv Epoc EEG headset was used to capture the brain activity and the open source ITU-gaze tracker, presented in section 3.3, was used to capture the gaze direction of the user.

In a concert setting we wanted to avoid the machine learning approach. Instead we proposed an approach that gives continuous arousal-valence estimated values and no training of any classifier would be required. The estimated *Arousal* and *Valence* values were computed using the following formulas:

$$Arousal_{4s} = PBF3_{4s}/PAF3_{4s} + PBF4_{4s}/PAF4_{4s},$$

$$Valence_{4s} = PBF3_{4s}/PAF3_{4s} - PBF4_{4s}/PAF4_{4s},$$

, where $PBF3_{4s}$ is the beta power of EEG location F3 over the last 4 seconds, $PAF3_{4s}$ is the alpha power of EEG location F3 over the last 4 seconds, $PBF4_{4s}$ is the beta power of EEG location F4 over the last 4

seconds and $PAF4_{4s}$ is the alpha power of EEG location F4 over the last 4 seconds.

For the purpose of the performance these values were normalized over 20 seconds. The $Arousal_{20s}$ and $Valence_{20s}$ values were also computed over a window of 20 seconds. The normalized values were computed as follows:

$Arousal_{norm} = \frac{Arousal_{5s} - Arousal_{20s}}{Arousalmax_{norm}},$

$Valence_{norm} = \frac{Valence_{5s} - Valence_{20s}}{Valencemax_{norm}},$

, where $Arousal_{norm}$, $Valence_{norm}$ are the normalized arousal and valence values and $Arousalmax_{norm}$, $Valencemax_{norm}$ are the maximum absolute values acquired when subtracting the 20s values from the 5s values for all sent values from the beginning of the performance.

The normalized values were then sent to the EyeHarp DMI through the VRPN (Virtual-Reality Peripheral Network) protocol.

The EyeHarp DMI an arpeggiator is implemented as described by Vamvakousis and Ramirez (2011). Four different arpeggios were generated, of varying speed. When high arousal values were received ($Arousal_{norm} > 1$), the volume of the fast arpeggios was increased, while the volume of the slow arpeggios was decreased. The opposite happened for low arousal values. This mapping was chosen considering that fast tempos reflect more emotional tension in music performance , whereas slow tempos are related to more relaxed emotional states (van der Zwaag et al. (2011)). Arousal values also affected the brightness of the "pie" in the interface.

Depending on the estimated valence value, three possible chord sequences were triggered. For low valence values ($Valence_{norm} < -0.2$) (sad or angry emotions), the triggered chord sequence was: Am, Dm, Em, Am. For average values ($-0.2 < Valence_{norm} < 0.2$), the triggered sequence was: Am, F, G, Am. Finally for positive valence ($Valence_{norm} > 0.2$), the sequence C, F, G, C was triggered. Additionally, the estimated valence affected the color of the "pie" in the interface. Low valence resulted to green variations, while high valence to red variations.

4.5   Event-Related Potential-based Interfaces

Over the past two decades BCI research has explored a variety of approaches for collecting, analyzing, and interacting with brain activity data. In most
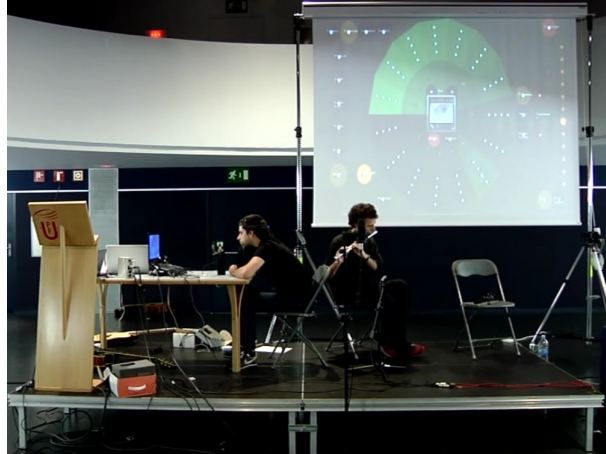
Figure 4.8: EyeHarp performance with EEG emotion recognition. A video of the performance can be found at https://youtu.be/Wc9690XKmuE

cases, the information is encoded voluntarily by the user, either by performing some mental task producing a measurable signal to be used as a command, or by selectively attending to one of the presented stimuli to encode a choice. Selective attention is often detected by observing event related potentials (ERPs), in particular the P300 wave whose occurrence is related to the persons reaction to a particular stimulus, and not to the physical attributes of the stimulus. P300 potentials, when recorded by electroencephalography (EEG), can be observed as a positive deflection in voltage with a latency (i.e. delay between the stimulus and the response) of roughly 250-500 milliseconds (Sellers and Donchin (2006); Polich (2007)). They are usually elicited using the oddball paradigm, in which low probability target stimuli are randomly mixed with high probability non-target ones.

In the past, visual P300 responses have been widely investigated for implementing BCIs (e.g. Bayliss et al. (2004); Piccione et al. (2006), and in particular for creating speller applications (Sellers and Donchin (2006); Farwell and Donchin (1988); Lenhardt et al. (2008); Nijboer et al. (2008a)). Similarly, auditory P300 responses have been used for implementing speller applications. In a study conducted by Furdea et al. (2009), a matrix of characters is presented for reference purposes with its columns and rows marked by a spoken number that is presented to the subject. Subjects are instructed to attend to the spoken number, which identifies the character.

When the spoken number corresponding to the row or column containing the character is produced, it elicits a P300 wave, which can be detected from the EEG. The selected letter is identified according to the row and column that give a P300 response. The evaluation of the system produced satisfactory results with performance reaching up to 100% for one subject. However, it is clear that auditory stimulation with spoken numbers is time consuming, reducing the information transfer rate (selection of a letter can take 3.6 minutes).

In a more recent study conducted by Klobassa et al. (2009), the spoken numbers were replaced by 6 natural sounds, which were mapped to rows and columns in an intuitive way allowing subjects to learn the mapping within a couple of sessions. Subjects were divided into two groups: one group was given auditory and visual stimulations while the other received only auditory stimulation. Although at the beginning of the experiment the accuracy of the auditory-only group was lower than the accuracy of the auditory-visual group, after 11 sessions their accuracy increased comparable to the one of the auditory-visual group. Inter-Stimulus interval was 500 ms and the reported average ITR for the auditory modality was 1.86 bits/min.

Most oddball experiments use acoustic cues such as pitch, amplitude or length. However, other sound properties, such as spatial location of the stimulus, have been investigated. Halder et al. (2010), conducted a 3-class oddball experiment comparing 3 auditory modalities: pitch, direction and amplitude. The pitch modality gave better performance on 70% of the participants (average ITR 1.70 bits/min). Teder-Sälejärvi and Hillyard (1998), conducted an oddball experiment in which an array of seven speakers (with a separation among them of 9 degrees) presented targets and non-targets in random order. Subjects attention to a particular direction elicited P300 responses. In a study conducted by Schreuder et al. (2010), explored the use of virtual spatial localization to separate targets from non-targets through stereo headphones. Non-targets were produced from a straight direction (i.e. zero degrees) while targets were produced from a 30 and 90 degrees direction. The focus of this study was on early mismatch negativity potentials and not in P300 responses, engaging the subjects in passive listening while they were watching a film. A similar study (Deouell et al. (2006)) was conducted using free-field speakers with 10 degrees spatial separation.

In a more related study (Schreuder et al. (2010)), a multi-class BCI experiment, which used spatially distributed, auditory cues was conducted. The stimulus set consisted of 5 stimuli, different in pitch. The subjects were

surrounded by 5 free field speakers, each of which was assigned to one of the stimuli. In the experiment, 5 subjects participated in an oddball task with the spatial location of the stimuli being a discriminating cue. The experiment was conducted in free field, with an individual speaker for each location. Different Inter Onset Intervals (IOIs) were investigated: 1000, 300, and 175 ms. Average accuracies were over 90% for most conditions, with corresponding information transfer rates up to an average of 17.39 bits/minute for the 175 ms condition (best subject 25.20 bits/minute). Interestingly, when discarding the spatial cues by presenting the stimuli through a single speaker, selection accuracy dropped below 70% for most subjects.

In a later study (Schreuder et al. (2011), an auditory speller was implemented using the same stimuli presentation design as in Schreuder et al. (2010), but increasing the set to 6 sounds. In order to optimize the spelling speed, a dynamic stopping method was introduced. This method minimized the number of repetitions required for each trial. Sixteen out of 21 subjects managed to spell a sentence in the first session. These subjects were selected for a second session where they were asked to type two sentences. In the second session an average of 5.26 bits/min (0.94 chars/min) ITR was achieved, which sets the current state of the art in auditory P300 spellers.

A very similar auditory BCI system using spatially distributed auditory cues is proposed by Käthner et al. (2013). The set of free field speaker is replaced by stereo headphones. Different IOIs of 560, 400, 320, 240 and 160 ms were evaluated in a P300 auditory speller paradigm. An average of 2.76 bits/min was reported under the 400 ms IOI condition. Unfortunately the training of the classification process was performed only for the 560 ms IOI. The acquired classifier was then used for all studied IOIs. This resulted in the biased conclusion that bigger IOIs give better selection accuracy. The opposite results were obtained by Schreuder et al. (2010), when a separate classifier was trained for each condition.

In another study Höhne et al. (2011), a 9-class auditory ERP paradigm was proposed where stimuli varied in pitch and stereo panning. Three stimuli different in pitch (high, medium, low) were presented from three different locations (left, center, right) using headphones. The proposed auditory ERP paradigm was used in a predictive text system. Users were able to spell on average 0.8 characters per minute (3.4 bits/min). In follow-up study conducted by Höhne et al. (2012), the artificially generated sounds were replaced with spoken syllables and the two conditions were compared. Improved classification performance and ergonomics were observed when

using spoken syllables as stimuli. In a more recent study conducted by Simon et al. (2014), a stimuli set consisting of 5 animal tones was used in an auditory speller BCI. The reported average online ITR across 11 users was 3.29 bits/min.

Other researchers have investigated the feasibility of using the Emotiv EPOC device for detecting auditory ERPs. Badcock et al. (2013) simultaneously recorded, using research and Emotiv Epoc devices, the EEG of 21 subjects while they were presented with 566 standard (1000 Hz) and 100 deviant (1200 Hz) tones under passive and active conditions. For each subject, they calculated auditory ERPs (P1, N1, P2, N2, and P300 peaks) as well as mismatch negativity (MMN) in both active and passive listening conditions. They restricted their analysis to frontal electrodes. Their results show that the morphology of the research and Emotiv Epoc EEG systems late auditory ERP waveforms were similar across all participants, but that the research and gaming EEG system MMN waveforms were only similar for participants with non-noisy MMN waveforms. Peak amplitude and latency measures revealed no significant differences between the size or the timing of the auditory P1, N1, P2, N2, P3, and MMN peaks. Based on these results it was concluded that the Emotiv Epoc EEG system may be a valid alternative to research EEG systems for recording reliable auditory ERPs. These results are confirmed by another recent study conducted by Wang et al. (2015). In another study conducted by Duvinage et al. (2013), Emotiv EPOC was compared against the ANT device in a 4-class visual oddball task. Emotiv underperformed ANT (86% versus 93% selection accuracy) among 9 healthy subjects under seating conditions, concluding that although the Signal to Noise Ratio is lower in the Emotiv device, it could be effectively used in non-critical BCIs as a user-friendly, low-cost alternative. In another recent study conducted by Nijboer et al. (2015), Emotivs usability is compared with a 32-channel wet-electrode Biosemi system and an 8-channel dry electrode g.Sahara system. It is concluded that water-based or dry headsets should be used when aesthetics, easy setup and fun are important, while for applications that require high accuracy and efficiency gelled headsets should be preferred.

In another study conducted by Nijboer et al. (2015), Emotiv Epocs amplifier was combined with a standard infra-cerebral electrode cap with Ag/AgCl electrodes. The result was a low-cost portable EEG system that was tested in a single trial 2-class auditory oddball paradigm under sitting and walking conditions. With an IOI of 1 second, the single trial accuracy was 77% for sitting and 69% for walking conditions. The same proposed EEG system,

in a more recent study [28] was evaluated in a single trial 3-class auditory oddball task, under walking and sitting conditions. The mean classification accuracy was 71% for the seating and 64% for the walking condition.

4.6 Evaluation of low cost Emotiv headset for auditory ERP-based applications

Our research on Brain Computer Interfaces (BCIs) was conducted mainly using the Emotiv EPOC 14-channel wireless EEG headset. Emotiv Epoc is a low-cost EEG device, marketed mainly for gaming purposes. For that reason we were concerned about the credibility of the device in Brain-Computer applications. It consists of 16 wet saline electrodes, providing 14 EEG channels, and a wireless amplifier (with a sample rate of 128 Hz). The 16 electrodes are aligned with positions in the 10-20 system: AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, FC4, M1, and M2. The electrode positioned at M1 acts as reference electrode, while the electrode at M2 is used for reducing external electrical interference.

A good way to measure the quality of an EEG device is to compare its performance in standard BCI scenarios. In a study involving 10 subjects without any disability, we measured the performance of Emotiv headset in an auditory-oddball paradigm. We then compare its performance in terms of Information-Transfer Rate (ITR) with similar studies, in which expensive equipment is utilized. The information transfer rate (ITR) ((Pierce (2012)), i.e. the amount of information carried by every selection, can be computed as follows:

$$ITS(bits/min) = S \cdot \left[\log_2(N) + P \cdot \log_2(P) + (1 - P) \cdot log_2 \left(\frac{1 - P}{N - 1} \right) \right]$$

$$(4.2)$$

, where ITR is the number of bits per minute, S represents the number of selections per minute, N represents the number of possible targets, and P represents the probability that they are correctly classified. Note that increasing S by decreasing the number of repetitions would not necessarily increase the ITR because the accuracy of the classifier (i.e. P) will decrease. Thus, there is a tradeoff between S and P, and the choice of which one of the two is more important depends on the type of BCI application.

In a study involving 10 healthy subjects, we evaluate a 6-class auditory P300 paradigm in which the stimuli set consists of sounds of musical instruments

with different pitch, timbre and spatial distribution (i.e. panning). To the best of our knowledge, this is the first auditory oddball paradigm study which explores stimuli varying simultaneously in pitch, stereo spatialization and timbre.

4.6.1 Methods

The study enrolled 10 volunteer participants (six women) with normal hearing. Age ranged from 25 to 38 (mean = 30, SD = 4). Subjects reported to have normal hearing, and no difficulty with spatial localization of sounds in everyday situations. None of the participants had previously participated in a BCI study.

The Emotiv EPOC EEG system [3] was used for acquiring the EEG data. It consists of 16 wet saline electrodes, providing 14 EEG channels, and a wireless amplifier (with a sample rate of 128 Hz). The 16 electrodes are aligned with positions in the 10-20 system: AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, FC4, M1, and M2. The electrode positioned at M1 acts as reference electrode, while the electrode at M2 is used for reducing external electrical interferences. We collected and processed the data using the OpenViBE platform (Renard et al. (2010). In order to trigger virtual instrument sounds through the OpenVibe platform, a VRPN to midi gateway was implemented and used along with LoopBe virtual MIDI port . Sound stimulus was then generated by Propellerhead Reason virtual instrument host application. Both data acquisition and on-line scenario were performed on a laptop with an Intel Core i5 2,53 Ghz processor with 4 GB of RAM, running windows 7 64-bit Operating System and using the laptops internal sound card (Realtek ALC269). By analyzing the audio recording of a session, the overall output audio latency was found to be 46 ms and the standard deviation 4.38 ms.

Subjects were sat motionless in a comfortable chair facing two loudspeakers, Roland MA-150U placed at 45 and -45 degrees with respect to the subject's orientation. The speakers were placed 15 cm below ear level and approximately at one meter from the subject (see figure 4.9). At the beginning of each experiment, subjects were asked to maintain their eyes closed, avoid any facial or body movement. The room was not electromagnetically shielded, and no extensive sound attenuating precautions were taken. All the experiments were designed as an auditory oddball task.

---

[3]http://emotiv.com/epoc/
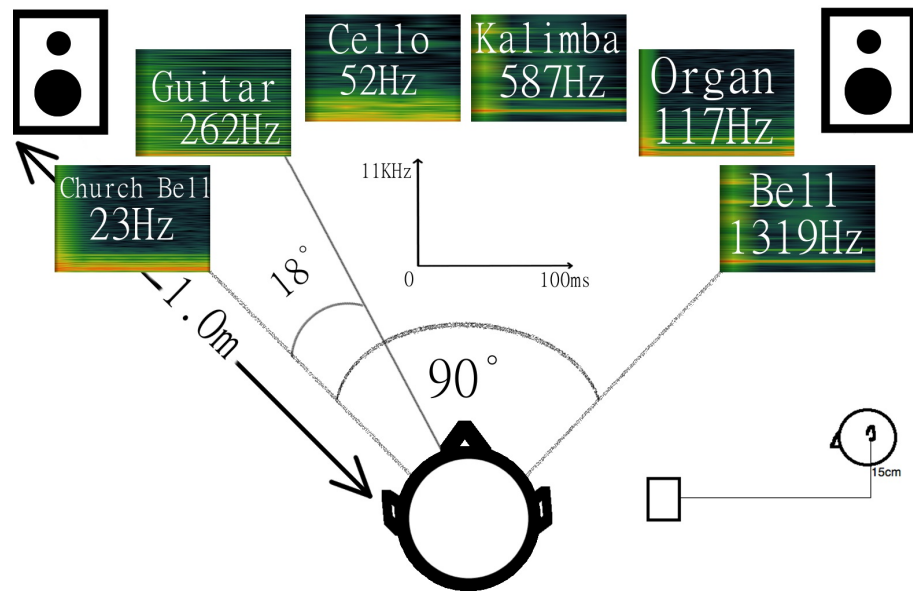
Figure 4.9: The sound stimuli are homogeneously distributed in the stereophonic space. There is a steady 9th musical interval between two consecutive sounds in pitch. The spectrograms of all stimuli are also shown.

The nature of the auditory stimuli has an effect on the performance of auditory ERP-based BCIs. In the current study the stimuli set varied in 3 aspects: timbre, pitch and stereo panning. The sound stimuli set consisted of 6 musical instrument sounds: church bell, guitar, cello, kaliba, organ and balinese bell. The stimuli were homogeneously distributed in the stereophonic space and varied in pitch (see figure 4.9). The lowest pitch corresponded to the church bell (F#0 / 23.12Hz). followed by the cello (G#1 / 51.91Hz), the organ (A#2 / 116.54Hz), the guitar (C4 / 261.63Hz), the kalimba (D5 / 587.33Hz) and the balinese bell (E6 / 1318.51Hz).

The musical interval between two consecutive pitches was a major 9th. The stimuli were placed in the stereophonic space so as to avoid having sounds with similar timbre or pitch close to each other. The speakers were set to equal loudness intensity of 60 dB SPL for every stimulus.

While setting up the EEG device, subjects were exposed to each stimulus in isolation and then to the stimuli mix in order for them to be familiarized with the sounds. This procedure lasted 5-10 minutes. Then for each subject, a training session was followed by an online session. A training session
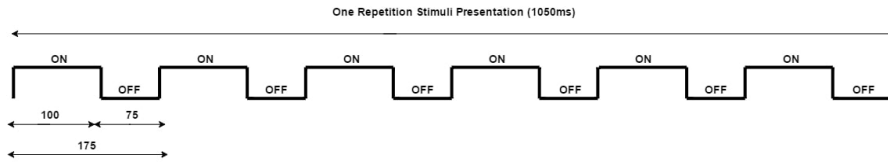
Figure 4.10: Stimuli presentation during a repetition

consisted of 10 trials while an online session consisted of 20 trials. Each trial consisted of 25 repetitions. Each repetition consisted of one presentation of each of the six sound stimuli in a pseudo-random order, with the constraint that during a trial two occurrences of the same stimulus were separated by at least two other stimuli. The Inter Onset Interval (IOI) was 175 ms and the duration of each stimulus was 100 ms (see figure 4.10). At the beginning of each trial a stimulus was randomly selected as a target and played back three times (with IOI 175 ms), followed by a pause of 500 ms before the trial started. The subjects were asked to count the occurrences of the target stimulus of each trial. For each subject 3750 non-target and 750 target epochs were recorded taking into account both the training and online sessions.

The recordings of the training sessions were analyzed in order to obtain a spatial filter and train a two class (target, non-target) LDA classifier (see figure 4.11). The EEG signals were downsampled to 32 Hz, digitized with a resolution of 16 bits, and band-pass filtered with a 4th order Butterworth 1-12Hz filter. Given the noisy nature of the EEG signal, a xDAWN spatial filter was applied in order to enhance the P300 response. The xDAWN algorithm, proposed by Rivet et al. (2009), allows the estimation of a set of spatial filters for optimizing the signal to signal-plus-noise (SSNR) ratio. The xDAWN method assumes that there exists a typical response synchronized with the target stimuli superimposed on an evoked response to all the stimuli, and that the evoked responses to target stimuli could be enhanced by spatial filtering. A window of 250 to 750 ms after the stimuli presentation was applied to train the xDAWN algorithm in order to obtain a 14 to 3 channels spatial filter. Given the mean audio latency of 46 ms, this window is equivalent to a window of 204 to 704 ms. This resulted in a matrix of 48 features. The features produced by the xDAWN filter were used to train a classifier f of the form:

$$f(Fs([t + 250, t + 750])) \rightarrow \{target, non - target\}$$
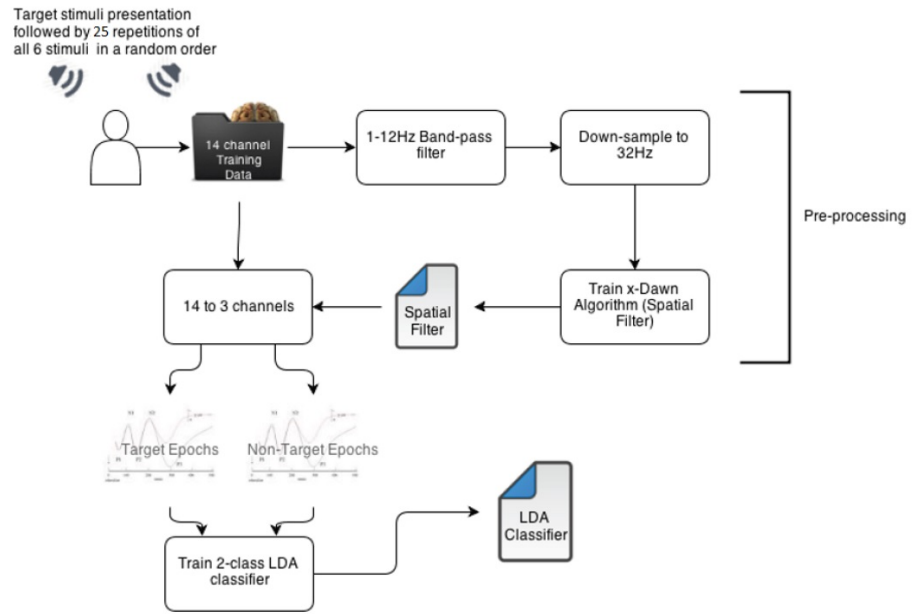
Figure 4.11: Obtaining a Spatial Filter and an LDA classifier.

,where t is the stimulus presentation time, Fs([t+250,t+750]) is the feature set generated by the spatial filter, and target and non-target are the classes to be discriminated. Classification was performed by applying linear discriminant analysis (LDA) to the training data. LDA finds a linear combination of features, which separates two or more classes of objects or events. The resulting combination may be used as a linear classifier.

During the online session, the 48 features vector for each epoch were fed to the obtained LDA classifier (see figure 4.12), whose output consisted of the vector distance to the hyper-plane (negative value for targets and positive for non-targets). These values were fed into a voting classifier. At the end of each trial (after 25 repetitions of each stimulus), the voting classifier sums up the hyper-plane distance of each stimulus. The stimulus with minimum sum is selected as the predicted target for that trial.

4.6.2   Results

Accuracy and ITR

We distinguish between two accuracy measures: classification and selection accuracy. Classification accuracy refers to the percentage of sub-trials that is correctly identified as target or non-target. Selection accuracy refers to the percentage of trials in which the target stimulus is correctly identified. Given that we are interested in detecting target stimuli, in the following we report on selection accuracy.

Table 4.13 presents the online accuracy of each subject. The minimum accuracy for having a practical BCI is considered to be 70% (Kübler et al. (2004)). Seven subjects exceeded this threshold. The optimum number of repetitions which maximize the ITR was also computed for each subject (offline ITR) in the following manner: for the 7 subjects that achieved maximum accuracy greater or equal to 70% the number of repetitions that maximizes the ITR was computed only for the repetitions that gave accuracy more than 70%. For the 3 subjects with less than 70% accuracy, we computed the number of repetitions required to reach the maximum accuracy.

It is interesting to point out that the best subject (F28) only needed 4 repetitions to achieve 75% accuracy, resulting in 17.05 bits/min offline ITR. The average accuracy across all subjects over different number of repetition is shown in figure 4.13. The average selection accuracy across all subjects increased as more repetitions were added. In some applications where accuracy is more important than ITR it would make sense to add more repetitions per trial.

Amplitude and latency of P300

The P300 potential is described by its amplitude (V) and latency (ms) within a certain time window. Here we define the P300 amplitude as the voltage difference (V) between the smallest negative and the largest positive peak (peak-to-peak measurement). Both the online and training session data were merged and analyzed using ERPLAB toolbox. The time windows used for measuring the positive and negative peak amplitude were 220 to 700 ms and 120 and 220 ms after stimulus onset, respectively. Latency was defined as the period between stimulus onset and the time when the maximum positive peak amplitude was reached. In the present task, the overall average largest difference between target and non-target across all subjects

	Online (25 repetitions)		Offline (#repetitions to maximize ITR)		
	Accuracy	ITR (bits/min)	Repetitions	Accuracy	ITR
F28	95%	4.99	4	75%	17.05
F25	85%	3.72	13	85%	7.15%
M38	75%	2.72	15	70%	3.84
F31	75%	2.72	25	75%	2.73
M35	70%	2.30	23	70%	2.5
M28	70%	2.30	23	70%	2.5
M32	65%	1.92	25	65%	1.92
M26	60%	1.57	25	60%	1.57
M30	55%	1.25	21	55%	1.49
AVG	72%	2.47	19,9	70%	4.3

Table 4.3: Online and offline accuracy for all subjects. In the case of the offline analysis the optimum number of repetitions that maximize the ITR was computed

was observed at location AF4. Table 4.4 shows the peak amplitudes and latencies for each subject at location AF4. Figure 4.14 shows the averaged target and non-target waveforms of each user at the locations with maximum target and non-target difference for each subject. The latency values of table 4.4 are corrected, considering the mean audio latency (46 ms).

	Target Epochs		Non-Target Epochs	
Subject	Latency (ms)	Peak (V)	Latency (ms)	Peak (V)
F28	579	6.10	438.38	2.67
F25	602.44	5.59	172.75	0.76
M38	243.06	3.12	414.94	1.85
F31	266.5	5.12	610.25	1.89
M31	641.5	4.27	297.75	1.78
M35	485.25	1.45	180.56	0.58
M28	461.81	3.02	641.5	2.25
M32	313.41	5.63	563.38	1.71
M26	563.38	7.95	204	3.93
M30	539.94	3.97	235.25	0.14
AVG	469.63	4.62	375.88	1.76

Table 4.4: Peak amplitudes and latency for each subject at location AF4

Sound Comparison

Table 4.5 shows the confusion matrix of all stimuli for all users. The diagonal in the table shows the number of correct selections (hits), whereas the other cells display the number of incorrect selections (errors).

	ChBell / F#0 / -45	Cello / G#1 / -9	Organ / A#2 / 27	Guitar / C3 / -27	Kalimba / D4 / 9	Bell / E5/ 45	Selection Accuracy(%)	N
ChBell	24	2	5	0	1	3	68.57%	35
Cello	1	22	1	3	4	4	62.86%	35
Organ	1	0	23	0	1	2	85.19%	27
Guitar	2	0	0	24	1	5	75.00%	32
Kalimba	1	1	4	3	23	1	69.70%	33
Bell	3	2	2	1	2	28	73.68%	38
PPV	75.00%	81.48%	65.71%	77.42%	71.88%	65.12%		

Table 4.5: Confusion matrix of classification results over all subjects. Selection accuracy, positive predictive value (PPV), and the number of trials where each stimulus was selected as a target (N) are given for convenience.

4.6.3 Discussion

Accuracy and ITR

In order to evaluate the performance of the proposed system we should compare with the state of the art auditory ERP-based BCIs. As each study performs a different auditory oddball task, probably the most indicative value in order to compare the performance of different studies is the achieved ITR. Table 4.6 shows among other, for each study, the best online and/or offline accuracy and the average ITR achieved.

The maximum ITR (17.39 bits/min) has been achieved by Schreuder et al. (2010) in a 5-class auditory oddball paradigm experiment involving 5 subjects. The stimuli set consisted of 5 artificially generated harmonic sounds, varying in pitch and 360 spatialization. Fifteen 15 repetitions were recorded per trial with 175 ms IOI. However, this study only reports the offline accuracy: the ITR is computed by taking into account only the minimum number of repetitions per trial, computed offline in order to achieve more than 70% accuracy and maximize the ITR. The second best offline ITR is reported by Furdea et al. (2009) in an speller paradigm experiment involving 13 subjects. The stimuli set consisted of 5 spoken numbers. Thirteen

	Discriminating Cues	Task	Number of Subjects	IOI (ms)	Average Selection Accuracy %	Average ITR (bits/min)	#Sequences	EEG system	Classifier
Schreuder et al. (2010)	Pitch, 360 Spatialization. (5 Speakers)	5 class oddball	5	175	94%	17.39 - Offline, ONS	15	Brain Products varying number of electrodes	Fisher Discriminant Analysis (FDA)
Schreuder et al. (2011)	Pitch, 360 Spatialization (5 Speakers)	hex-o speller	21	175	77.4 %	5.26 - Online. Dynamic Stopping Method.	15	Brain Products 56 electrodes	(Linear Discriminant Analysis) LDA
Höhne et al. (2011)	Pitch, Stereo Spatialization (Headphones)	9-class oddball driving a predictive text entry system	12	225	89.37%	3.4 - Online.	15	Brain Products, 63 electrodes	FDA
Simon et al. (2014)	Animal tones.	5x5 speller	11	400	76% Online 90% Offline	3.29 -Online 4.23 - Offline	10	Brain Products 28 electrodes	Stepwise (swLDA) LDA
Käthner et al. (2013)	Pitch, Stereo Spatialization (Headphones)	5x5 speller	20	400	65%	2.76 - Online	10	Brain Products 28 electrodes	Stepwise (swLDA) LDA
Höhne et al. (2012)	Pitch, Stereo Spatialization, Spoken syllables. (Headphones)	9-class oddball	9	130	not reported	5.31 - simulated online with Dynamic Stopping Method.	20	Brain Products, 63 electrodes	FDA
Current Study	Timbre, Pitch, Stereo Spatialization. (Speakers)	6-class oddball	10	175	72%	2.47 -Online 4.3 - Offline, ONS	25	Emotiv EPOC, 14 electrodes	LDA, xDawn
Halder et al. (2010)	Pitch (Headphones)	3-class oddball	20	480-600	78.5%	2.46 - Offline	2	Brain Products 67 electrodes	swLDA
Klobassa et al. (2009)	Timbre, Pitch. (Speakers)	6x6 matrix speller	10	500	64.07	1.86 - Online	8	SA Electronics 64 electrodes	swLDA
Furdea et al. (2009)	Speech (Headphones)	5x5 matrix speller	13	625	65%	1.54 -Online 4.66 - Offline, ONS	15	gtek 16 electrodes	swLDA
De Vos et al. (2014)	Pitch, Stereo Spatialization. (Headphones)	3-class oddball	20	1000	71%	1.07 - Online	1	Modified Emotiv	swLDA

Table 4.6: A comparison of existing auditory P300 BCI studies. (ONR = Optimal Number of Sequences to maximize ITR)

repetitions were recorded per trial with 625 ms IOI. After determining the minimum number repetitions to maximize the ITR, the reported ITR was 4.66 bits/min. Halder et al. obtained an ITR of 2.46 in an offline 3-class auditory oddball paradigm experiment involving 20 subjects. In our study, after determining the minimum number of repetitions per trial, the average ITR among all users was 4.3 bits/min. With the exception of the ITR obtained by Schreuder et al. (2010), the achieved ITR of our study is comparable with the ITR the rest of the offline auditory ERP studies.

The maximum online ITR for an auditory BCI (5.31 bits/min) is reported by Höhne et al. (2012) who simulated online sessions using pre-recorded data and applied a dynamic stopping method. The average ITR of the simulated sessions sets the state of the art of online auditory BCIs. However, these results refer only to simulated sessions, not to real online sessions. The maximum ITR achieved in an online experiment is reported in another study by Schreuder et al. (2011) in which a 6-class auditory oddball paradigm utilizing spatial and pitch discrimination cues was applied in hex-o speller. The reported average ITR was 5.26 bits/min. In this study a dynamic stopping method is applied, minimizing automatically the number of repetitions per trial and as a result improving the ITR. In another online study by where no dynamic stopping method was applied the obtained ITR was 3.4. They studied a 9-class oddball paradigm driving a predictive text entry system, which involved 12 participants. The online ITR results obtained in our study (2.47 bits/min) are comparable with the result reported by Höhne et al. (2011). Applying a dynamic stopping method in future applications might increase our achieved ITR.

The difference between the obtained online and offline ITR in the current study and those obtained by the state-of-the-art systems may be due to the quality of the EEG signal acquired by the different EEG devices and the different number of channels used. In the studies with the best online and offline ITRs reported (Schreuder et al. (2010, 2011)) a BrainAmp amplifiers with 56 or more Ag/AgCl electrodes was used. The EEG signals were sampled at 1 kHz and filtered by an analog bandpass filter between 0.1 and 250 Hz. After applying the analog filter, the data was low-pass filtered to 40 Hz and downsampled to 100Hz. In the case of the online P300 speller study (Schreuder et al. (2011)), the features used for classification consisted of two to four amplitude values per channel resulting in 112224 features. In the case of the 5-class oddball study (Schreuder et al. (2010)), the 20 channels that accounted for most of the difference between the two classes were automatically selected within each fold of the cross validation. The 10

channels with the highest positive ROC peak and those 10 with the lowest negative ROC peak were used. Data from these channels were decimated by taking the mean of five samples, reducing the data to 16 post-baseline samples per channel. Samples from all 20 channels were then concatenated to form a 320 dimensional feature vector. In the current study the Emotiv Epoc EEG device was used. Using 16 saline electrodes, it provides 14 channels in fixed positions and a sampling rate of 128 Hz. The signal was downsampled to 32 Hz and xDawn algorithm was applied to generate a 14 to 3 channels spatial filter. Epochs had a duration of 500 ms. As a result 48 features were fed to an LDA classifier. The number of channels and the sample rate of the of the Emotiv device is lower when compared to the above systems and its price (600$US for the research edition) is about 50 times lower. The difference in accuracy of the Emotiv device and the above systems has been also observed in visual ERP-based BCIs (Duvinage et al. (2013)). However, the ITR achieved in the current study is comparable to the obtained ITR by most of the reported state-of-the-art systems, showing that an auditory ERP-based BCI is feasible with low-cost EEG devices and an off-the-shelf laptop computer.

Amplitude and latency of P300

It is important to mention that the exact response to each stimulus was impossible to be analyzed in the current study, because of the small IOI. Due to the size of the considered IOI, the evoked potentials of different stimulus overlap. Nevertheless, target epochs were characterized by higher amplitudes and longer latency. The overlapping of evoked potentials of different stimulus is common in P300 systems where maximization of the ITR is important. As shown in Table 4.6, most of the systems mentioned in this section share this characteristic, with the exception of the systems proposed by Halder et al. (2010) and De Vos et al. (2014). These systems are 2-trial and single trial systems, respectively, and thus a large IOI has little impact in the resulting ITR.

Sound Comparison

Sound stimuli may vary in pitch, timbre, spatialization and combinations of those. Table 4.6 summarises the discriminating cues and auditory task studied in previous studies. Halder et al. (2010) compared the achieved ITR of stimuli varying in either pitch, spatialization or amplitude. It was concluded that the maximum mean ITR was achieved when stimuli varied

in pitch, followed by spatialization. In a later study, Schreuder et al. (2010) indicated that combining pitch and directional discrimination cues, provides much higher accuracy than applying just pitch (94% versus 56%). Höhne et al. (2011) also made use of these two discriminating cues, while in a later study Höhne et al. (2012) replaced the artificially generated sounds with spoken syllables. In a 5-class oddball paradigm. Klobassa et al. (2009) performed a 6x6 matrix auditory P300 speller study in which the stimuli varied in pitch and timbre (sounds of musical instruments). However, to the best of our knowledge there is no study in which timbre, pitch and spatialization are combined in a single auditory oddball paradigm.

In the current study all stimuli varied in all these 3 discriminating cues. The advantage of this approach might be that the task of distinguishing one stimulus from the others might be easier by providing more discriminating cues. Nevertheless this approach might lead to stimuli sets in which not all stimuli are equally salient. This might be the case for the Organ/A#2 stimulus in the present study. It was the stimulus with the highest selection accuracy, but lowest PPV. On the other hand the Cello/G#1 was the stimulus with the lowest selection accuracy and highest PPV. This suggests that the organ was the most salient stimulus while the cello was the least salient one. The most balanced stimulus was the guitar, as it gives both high accuracy and high PPV. Unbalanced -in terms of saliency- stimuli sets might harm the overall performance of auditory BCIs. For instance, a P300 speller could favor the selection of specific letters, but impede the selection of others. If the speller interface includes deleting and re-spelling misspelled characters, a user could spend considerable time trying to spell a single character. Using musical instrument sounds and three auditory discrimination cues makes the task of creating a balanced stimuli set a difficult task. Saliency of music instrument sounds with different timbre is subjective. One solution to this problem might be allowing the final users of the proposed interfaces experiment and select each of the stimuli themselves, in order to achieve a balanced set.

Stimuli with similar stereophonic panning or pitch were not observed to be confused with each other more than the average. This can be explained by the fact that neighbouring sounds were produced by instruments with pitches which differ from each other by a musical interval of a major 9th (14 semitones). In the study with highest ITR in auditory P300-based BCIs, by Schreuder et al. (2010), the selected musical interval between two consecutive stimuli was a major 2nd (2 semitones). However, Schreuder et al. use a different loud-speaker for each stimulus, i.e. each sound is generated from

a different loud-speaker, for implementing directional cues. Using stereo panning instead of real stimuli spatialization, produces a weaker directional cue. In the current study, this fact is compensated by generating stimuli with larger pitch intervals.

Summary and Conclusions

Commercial low-cost EEG devices released in the last few years are bringing Brain-Computer Interfaces (BCIs) out of the laboratory. In the present study we investigated the feasibility of an auditory P300 interface when using the low-cost EEG Emotiv EPOC device. Although there are a few studies indicating the capability of Emotiv system to measure auditory ERPs, to the best of our knowledge, this is the first time a non-modified Emotiv system is being evaluated in an auditory ERP-based BCI. A 6-class auditory oddball paradigm experiment was performed on 10 healthy subjects with online sessions consisting of 20 trials. When averaging over 25 repetitions, the mean online selection accuracy across all users was 72%, with one participant achieving 95% accuracy. The average online ITR among all subjects was 2.47 bits/min. These results confirm that the Emotiv EPOC is capable of capturing auditory evoked potentials as has been reported in previous studies (Badcock et al. (2013); Wang et al. (2015)) and that the proposed BCI system can provide a low-cost mean of communication for healthy subjects. Our results suggest that the Emotiv EPOC device together with off-the shelf computer equipment and a simple and portable set-up, may provide a promising low-cost technology that could be used in auditory ERP-based BCIs.

4.7 Exploring timbre, pitch and spatialization auditory cues of auditory ERP paradigms

4.7.1 Introduction

In this study, we explore different combinations of timbre, pitch and spatial auditory stimuli (TimPiSp: timbre-pitch-spatial, TimSp: timbre-spatial, and Timb: timbre-only) and three Inter-Stimulus Intervals (ISI) (150ms, 175ms and 300ms), and evaluate our system by conducting an oddball task on 7 healthy subjects.

4.7.2   Methods

Participants

Eighteen healthy adults (8 female, mean age 34 years) participated in several multi-class auditory oddball paradigm. Subjects reported to have normal hearing, and no difficulty with spatial localization of sounds in everyday situations.

The Emotiv EPOC EEG system was used for acquiring the EEG data. The EEG signals were sampled at 128 Hz, digitized with a resolution of 16 bits, and band-pass filter with a 4th order Butterworth 1-12Hz filter. We collected and processed the data using the OpenViBE platform (Renard et al. (2010)). In order to trigger virtual instrument sounds through the OpenVibe platform, a VRPN to midi gateway was implemented and used along with LoopBe virtual MIDI port . Sound stimulus was then played back by Propellerhead Reason virtual instrument host application. MBOX low-latency sound card was used, offering 17 ms output latency. The LoopBe MIDI port used introduced an additional latency of 1 to 3 ms. Both data acquisition and on-line scenario were performed on a laptop with an Intel Core i5 2,53 Ghz processor with 4 GB of RAM, running windows 7 64-bit Operating System.

The purpose of this auditory modality experiment was to investigate how the physical characteristics of the sound stimuli of an auditory oddball paradigm affect the accuracy of the task. Three different ISI were explored: 300 ms and 175 ms and 150 ms. For the 300 ms and 175 ms conditions three different stimuli discriminating cues were examined: timbre only (Timb), timbre and spatial (TimSp), and timbre, pitch and spatial (TimPiSp). For the 150ms condition only the TimPiSp modality was studied. In the Timb conditions, all stimuli were generated with different timbre from each other but with fixed pitch (130.81 Hz) and spatial location (center); in the TimSp conditions, stimuli were generated with different timbre and spatial location from each other but fixed pitch; in TimPiSp conditions all timbre, pitch and spatialization were differentiated (see Table 4.7.2). In total 7 different conditions were studied: TimPiSp-150ms ISI (TimPiSp150), TimPiSp-175ms ISI (TimPiSp175), TimPi-175ms ISI (TimPi175), Timb175 (Timb175), TimPiSp-300ms ISI (TimPiSp300), TimPi-300ms ISI (TimPi300), Timb175 (Timb300). In all conditions the stimulus set consisted of 6 short sounds (of a duration of 100ms). Blocks of the different conditions were mixed to prevent time biases.

Timb		TimSp			TimPiSp		
Pitch (Hz)	**Stimuli**	**Pitch (Hz)**	**Stimuli**	**Spatial**	**Pitch (Hz)**	**Stimuli**	**Spatial**
130.81	Bell	130.81	Bell	-45	23.123	Bell	-45
130.81	Snare Drum	130.81	Snare Drum	-27	51.91	Cello	-27
130.81	Hi Hat	130.81	Hi Hat	-9	116.541	Organ	-9
130.81	Guitar	130.81	Guitar	9	261.626	Guitar	9
130.81	Kalimba	130.81	Kalimba	27	587.330	Kalimba	27
130.81	Claps	130.81	Claps	45	1318.51	Bali bell	45

Table 4.7: Cue properties in the different conditions

In all sessions, subjects were asked to sit motionless in a comfortable chair facing two loudspeakers, Roland MA-150U placed at 45 and -45 degrees with respect to the subjects orientation. The speakers were placed 15cm below ear level and approximately at one meter from the subject (see Figure 4.15). The speakers were set to equal loudness intensity of approximately 60 dB for every stimulus. Subjects were initially exposed to each stimulus in isolation and then to the stimuli mix in order to familiarize them with the sounds. At the beginning of each experiment, subjects were asked to close their eyes, minimize their eye movements and avoid moving during the experiment. The room was not electromagnetically shielded, and no extensive sound attenuating precautions were taken. All the experiments were designed as an auditory oddball task.

For each condition, a training session was followed by an online session. This resulted in 14 sessions for every subject, each one lasting 5-6 minutes, depending on the ISI. One minute break interfered between all sessions, and a 15 minutes break at the middle of the end of the 8th session. The whole experiment lasted about 2 hours per subject. The collected EEG data of each training session were used for training the xDawn algorithm for acquiring a spatial filter and a Linear Discriminant Analysis Classifier, used in the on-line classification process. Both the training and the on-line sessions consisted of ten trials. In the 300ms condition each trial consisted of 90 sub-trials, 15 for each stimuli, while in the 175 and 150ms conditions each trial consisted of 150 sub-trials, 25 for each stimulus. This resulted in 900 sub-trials per session (150 of which target) in the 300ms condition and 1500 sub-trials per session (250 of which target) in the 175 and 150ms conditions.

Before each trial a random stimulus was randomly selected as the target stimulus and was played back to the subject (see figure 4.16). A trial can be divided into N repetitions (where N is 15 for the 300ms conditions and 25 for

the 175 and 150ms conditions). A repetition consists of a random sequence of all 6 stimuli. An example of a repetitions stimuli presentation for the TimPiSp175 condition is shown in figure 4.17. Stimuli were randomized in a way that the same stimulus never appeared consecutively. The subjects were instructed to slightly move their index finger every time the target stimulus appeared and mentally count its occurrences. In the on-line session, 1.9 seconds after each trial, the stimulus detected as target was played back to the subject followed by an interval of 3 seconds before presenting the target stimulus of the next trial. Apart from the presentation of the detected target stimulus at the end of each trial, the training and online sessions were identical.

During the training session the eeg data was processed as explained in section 4.6.1 in order to acquire a spatial filter and an LDA classifier (see figure 4.11).

During the online session, the 48 features vector for each epoch were fed to the obtained LDA classifier (figure 4.12), whose output consisted of the vector distance to the hyper-plane (negative value for targets and positive for non-targets). These values were fed into a voting classifier. When the corresponding number of repetitions is reached, the voting classifier sums up the hyper-plane distances for all the repetitions of each stimuli. The stimulus with minimum sum is selected as the predicted target for that trial.

4.7.3 Results

ITR and accuracy

We distinguish between two accuracy measures: classification and selection accuracy. Classification accuracy refers to the percentage of sub-trials that is correctly identified as target or non-target. Selection accuracy refers to the percentage of trials in which the target stimulus is correctly identified. Given that we are interested in detecting target stimuli, in the following we report on selection accuracy. In order to investigate the system's accuracy for different number of repetitions, the voting classifier object in OpenVibe platform was modified to keep a log of the hyper-plane distances sums of each stimulus for any number of repetitions.

Tables 4.8, 4.9 and 4.10 provide the online accuracy of all subjects and conditions along with the number of repetitions in the on-line sessions. Figure 4.18 shows the average accuracy and ITR (among subjects) for different

number of repetitions. The ITR is considered to be zero, if the average
accuracy is less than 70%.

The maximum accuracy is found in the TimPiSp175 condition (97.1%), fol-
lowed by the TimPiSp150 (92.86%), TimbSp 175 (91.4%), Timb175 (88.57%),
TimPiSp300 (88.57%), TimbSp300 (84.3%) and Timb300 condition (80%).

The average accuracy exceeds 70% in all conditions after 10 repetitions
and 80% after 15 repetitions, while after around 18 repetitions the online
accuracy does not improve significantly in all conditions (see figure 4.18).
For a given number of repetitions, the 300 ms condition does not seem
to provide better accuracy than the 175 ms and 150 ms conditions and
as a result gives lower ITR. The maximum average ITR is achieved with
around 10-15 repetitions for all conditions. In the TimPiSp175 condition
the average accuracy is more than 90% after 19 repetitions. The maximum
average ITR is found In the TimPiSp150 condition (14.85 bits/min, with
an average of 9.43 iterations). The best subjects performance was in the
Timb175 condition (39.96 bits/min, accuracy 80% with 2 repetitions).

Physiological Response

For each condition and every subject, the training and on-line session EEG
recordings were merged into one dataset and analyzed in Matlab using
EEGlab (Delorme and Makeig (2004)) and ERPlab toolbox (Lopez-Calderon
and Luck (2014). This resulted in 3000 sub-trials (500 targets) for the
300 ms modality and 1800 sub-trials (300 targets) for the 175 and 150 ms
modalities, for each subject and condition. A window of 200 ms before the
stimulus presentation was used for baseline removal. In all conditions a
threshold of 150V was used for rejecting epochs with artifacts. The per-
centage of rejected epochs for each condition is shown in tables 4.8, 4.9 and
4.10. Since during the experiment, subjects remained still and with their
eyes closed, the high artifact rejection rate between sessions (raging from
0% to 74.4% for the same user) is due to noise introduced by the Emotiv
Epoch. Although the signal was always checked before every session, some
EEG channels became noisy in the middle of a session.

Initially a grand average for all 7 conditions was created for each subject,
and its P300 peak amplitude in the interval 250 and 650 ms was computed
for all EEG channels for the target epochs. For each subject, the EEG
channel with the highest P300 peak values was selected for further analysis.
Tables 4.8, 4.9 and 4.10 show the averaged P300 amplitude and latency for
all conditions and users. Figures 4.19 and 4.20 show the averaged target

Subject	300 ms ISI						175 ms ISI						150 ms ISI					
	Sel. Accuracy (%)	max ITR (bits/min) 70%	amplitude (V)	Latency (ms)	Electrode	Rejected Epochs (%)	Sel. Accuracy (%)	max ITR (bits/min) 70%	amplitude (V)	Latency (ms)	Electrode	Rejected Epochs (%)	Sel. Accuracy (%)	max ITR (bits/min) 70%	amplitude (V)	Latency (ms)	Electrode	Rejected Epochs (%)
M30	100 (7)	23.3 (3)	7.28	479	F8	4.2	100 (12)	15.9 (5)	5.24	639	F8	3.8	100 (11)	13.32 (7)	3.55	427	F8	1.2
M46	80 (13)	3,59 (13)	0.13	659	F4	0.2	90 (21)	11.5 (5)	2.36	478	F4	2.1	100 (23)	7.85 (16)	3.64	484	F4	0.7
M36	70 (8)	4.2 (8)	6.95	458	F4	0.9	100 (14)	13.4 (8)	6.99	505	F4	1.4	100 (10)	15.54 (6)	6.66	378	F4	41.5
M28	100 (15)	5,83 (8)	0,079	613	AF4	2.9	100 (25)	5,75 (10)	2.76	597	AF4	2.9	90 (24)	15.54 (6)	4.02	559	AF4	4.2
F28	100 (15)	8,97 (7)	3.53	433	F4	2.7	100 (17)	13.4 (8)	3.14	615	F4	1.9	100 (15)	11.5 (15)	3.07	648	F4	6.3
F58	90 (15)	4.8 (7)	5.18	436	F3	22.6	100 (25)	5,91 (25)	1.79	449	F3	6.9	70 (14)	6.66 (14)	1.42	391	F3	20.4
F57	80 (13)	3,59 (13)	1.25	503	F7	13.4	90 (20)	5 (16)	2.3	248	F7	1.6	100 (23)	33.57 (2)	2.09	267	F7	2.7
Mean	88.5 (12.3)	7.75 (8.4)	3.49	511		6.7	97.1 (19.1)	10.1 (11)	3.51	504		2.94	92.86 (17.1)	14.85 (9.43)	3.49	450.57		11.00

Table 4.8: Results for Timbre Pitch Spatial (TimPiSp) modality. For each condition and each user is given: (i) the Selection Accuracy and in parenthesis the Number of Repetitions Required, (ii) the Maximum ITR achieved and in parenthesis the Number of Repetitions that maximize it, under the constraint that at least a 70% of accuracy is achieved, (iii) the Amplitude in V and (iv) the Latency in ms of the P300 peak in the (v) given position and finally (vi) the percentage of rejected epochs during the off-line analysis.

| Subject | 300 ms ISI | | | | | | 175 ms ISI | | | | | |
	Sel. Accuracy (%)	max ITR (bits/min) 70%	amplitude (V)	Latency (ms)	Electrode	Rejected Epochs (%)	Sel. Accuracy (%)	max ITR (bits/min) 70%	amplitude (V)	Latency (ms)	Electrode	Rejected Epochs (%)
m30	100 (4)	33.5 (1)	6.61	519	F8	11.9	100 (9)	28.7 (2)	4.8	606	F8	1.6
m46	100 (7)	16.7 (2)	7.4	426	F4	0.1	90 (20)	6.15 (13)	5.13	466	F4	0.5
m36	60 (13)	-	9.03	369	F4	74.4	100 (9)	15.98 (5)	7.58	388	F4	0.5
m28	100 (6)	14.36 (6)	2.37	602	AF4	2.6	100 (7)	15.98 (5)	2.22	621	AF4	18.3
f28	80 (13)	3.59 (13)	4.83	443	F4	3.3	100 (17)	26.64 (3)	4.1	441	F4	0.7
f58	60 (13)	-	1.83	390	F3	12.7	80 (23)	3.47 (23)	2.77	296	F3	40
f57	90 (15)	4.8 (7)	1.57	245	F7	0.8	70 (10)	5.75 (10)	3	258	F7	2.5
Mean	84.3 (10.1)	10.4 (4.1)	4.8	428		15.1	91.4 (13.6)	14.7 (8.71)	4.2	439		9.16

Table 4.9: Results for Timbre Spatial modality (TimSp). Fields as in table 4.8. The ITR is not computed when the limit of 70% accuracy is not reached.

| Subject | 300 ms ISI | | | | | | 175 ms ISI | | | | | |
	Sel. Accuracy (%)	max ITR (bits/min) 70%	amplitude (V)	Latency (ms)	Electrode	Rejected Epochs (%)	Sel. Accuracy (%)	max ITR (bits/min) 70%	amplitude (V)	Latency (ms)	Electrode	Rejected Epochs (%)
m30	100 (5)	20,93 (3)	6,29	589	F8	0,4	100 (24)	9.99 (8)	0,19	649	F8	3,2
m46	70 (12)	2,8 (12)	5,45	546	F4	5,9	70 (14)	4.11 (14)	4,48	421	F4	7,8
m36	60 (14)	-	4,6	403	F4	4,9	100 (5)	39.96 (2)	8,62	341	F4	0
m28	80 (7)	6,66 (7)	2,05	551	AF4	1,5	90 (17)	9.99 (8)	2,59	559	AF4	1,2
f28	90 (14)	6,71 (5)	4,17	415	F4	4,7	90 (14)	19.18 (3)	5,22	382	F4	1,1
f58	80 (7)	6,66 (7)	8,37	392	F3	28,9	90 (17)	9.99 (8)	8,9	641	F3	15,9
f57	80 (10)	4,8 (7)	1,84	316	F7	1,5	80 (20)	4,43 (13)	2,44	409	F7	1,8
Mean	80 (9.9)	6,94 (5.86)	4,68	458,86		6,83	88.57 (15.86)	13.95 (8)	4,63	486		4,43

Table 4.10: Results for the Timbre modality. Fields as in table 4.9

and non-target responses of each users selected channel for all the 175 and 300 ms ISI conditions, respectively. In all plots, the red line corresponds to target epochs and the black line to non-target epochs. periodicity of 175 ms can be observed in the 175 ms condition and a periodicity of 300 ms in the case of 300 ms condition. As expected, this periodicity aligns with the stimuli presentation periodicity (see figure 4.17).

Figures 4.21, 4.22 and 4.23 show the average of all users target and non-target responses for all 300ms ISI conditions of 10 EEG channels. When comparing the 3 modalities, it is observed that while the target ERP responses are equally strong in all modalities, the TimPiSp gives the weakest non-target ERP responses, followed by the TimPi and the Timb modalities. This results in a stronger mismatch negativity value. This is also reflected in the selection accuracies of each of these modalities: 88.5%, 84.3% and 80% for the TimPiSp, TimSp and Timb modality respectively.

4.7.4   Discussion

For the first time the significance -in an auditory P300 paradigm- of the 3 most important perceptual auditory discriminating cues is studied: Timbre, Pitch and Spatialization, under three possible ISI conditions (300, 175 and 150 ms). The results of our study indicate that the best results are given when the stimuli are different in all three perceptual modalities, while shorter ISI results in higher ITR.

As seen in figures 4.19 and 4.20, all subjects have clear EPR responses in both the 175 and 300 ms conditions, although they vary in intensity and shape. The mean latency of the P300 peak for all 7 conditions is 468 ms, while no significant differences in the P300 peak amplitude and latencies are observed between the different conditions (see tables 4.8, 4.9, 4.10). Although the signal quality was checked at the beginning of each session, high epoch rejection rate was observed in some sessions. This might be due to the unstable behavior of saline water electrodes.

The channels with the strongest average P300 peak for all conditions were located in the frontal area for all subjects. When looking at the occipital channels though (figures 4.21, 4.22, 4.23), we can see an early positive deflection about 220 ms after the target stimuli presentation. This aligns with the results of Schreuder et al [7], where it is concluded that in the short 175 ms condition class difference has shifted toward the frontal areas when compared to the longer 1000 ms ISI condition.

Despite using a low cost EEG device, the performance of the proposed
system is comparable to state-of-the-art performance. In the TimPiSp150
condition the average selection accuracy obtained is 92.86% with 17.1 rep-
etitions and the average ITR is 14.85 bits/min with 9.43 repetitions. These
results compare well with the state-of-the-art results reported by Schreuder
et al. (2010) (selection accuracy 94%, with 11.6 repetitions; maximum ITR
of 17.39 bits/min, with 5.6 repetitions, PitchSpatial 175ms ISI).

The maximum average selection accuracy is found in the TimPiSp175 con-
dition (97.1%), followed by the TimPiSp150 (92.86%), TimbSp175 (91.4%),
Timb175 (88.57%), TimPiSp300 (88.57%), TimbSp300 (84.3%) and Timb300
condition (80%). The 300ms ISI conditions though were studied for a maxi-
mum of 15 repetitions, while the 175 and 150 ms ISI conditions were studied
for a maximum of 25 repetitions. Looking at figure 4.18, we can see that
for the same number of repetitions, the average accuracy is close for the
300 and 175 ms ISI conditions. The ITR though is much lower in the case
of 300ms ISI conditions, as more time is required for the same number of
repetitions. Thus, it is concluded that there is no reason for using long
ISIs in auditory P300 based BCIs. In order to get a significantly stronger
P300 response, When comparing the TimPiSp175 with TimPiSp150 con-
ditions, we see that although the first one gives better selection accuracy
(97.1% versus 92.86%), the second one achieves higher ITR (14.85 versus
10.1 bits/min). In the future, the ISIs limits should be studied in order to
determine the minimum ISI to maximize ITR.

In both 300 and 175 ms ISI conditions, the order of the conditions in terms
of selection accuracy is: TimPiSp, TimSp, Timb. Thus, it is clear that the
performance of the system improves as more discriminating cues are added.
This is also concluded when observing the averaged ERP responses of these
conditions (figure 4.23). Although the target stimuli responses have the
same intensity in all conditions, the non-target stimuli responses become
weaker as more modalities are added in the stimuli design. This results in
higher mismatch negativity values and thus, higher selection accuracy.

Schreuder et al. emphasized the importance of sound spatialization in stim-
uli presentation. However, in this case stimuli differed only in pitch and spa-
tialization. Selection scores went down below 70% for most subjects when
the spatialization modality was removed. Our results imply that when
stimuli are different in timbre, the spatialization still affects the selection
accuracy, but not so drastically. In the 300ms ISI conditions, the average
accuracy of TimSp modality is 84.3% while in the Timbre modality the ac-

curacy is 80%. In the 175 ms ISI conditions, the average accuracy of TimSp modality is 91.4% and the accuracy of the Timbre modality is 88.57%.

We have presented a multi-class BCI system based solely on auditory stimuli, which makes use of low-cost EEG technology. We have explored timbre-pitch-spatial, timbre-spatial, and timbre-only combinations of timbre, pitch and spatial auditory stimuli and three inter-stimuli intervals (150ms, 175ms and 300ms). We evaluated the system by conducting an oddball task on 7 healthy subjects. The maximum accuracy is found in the TimbPiSp175 condition (97.1%), followed by the TimPiSp150 condition (92.86%), TimbSp175 condition (91.4%), Timb175 condition (88.57%), TimPiSp300 condition (88.57%), TimbSp300 condition (84.3%) and Timb300 condition (80%). The maximum average ITR is found in the 150ms ISI, TimPiSp condition (14.85 bits/min, with 9.43 iterations). Lower Inter-Stimulus Intervals lead to higher ITR, while as more discriminating cues are added the selection accuracy and ITR increases. Based on the TimPiSp modality, an auditory P300 speller was implemented and evaluated by asking users to type a 12-characters-long phrase. Six out of 7 users completed the task. The average spelling speed was 0.56 chars/min and best subjects performance was 0.84 chars/min.

In this study we made use of an EEG device which is valued at about 50-100 times less costly than medical/research quality devices. However, interestingly our results are comparable to those achieved by medical devices. The obtained results show that the proposed auditory BCI is successful with healthy subjects and may constitute the basis for future implementations of more practical and affordable P300-based BCI systems. However, the high amount of noise introduced during some of the sessions (high epoch rejection rate in off-line analysis) affects the accuracy of the system, and thus for crucial BCI applications a more robust and stable EEG device should be used.

4.8 ERP-based BCMIs

Probably the most robust way of building a voluntarily controlled BCI that wouldn't require almost any training on behalf of the user, is through the P300 potential. The P300 potential is a positive deflection of the captured electromagnetic activity, 300ms after a rare or unexpected event is perceived, centred around the vertex of the cortex and spread all over the cortex. In a multi-class P300-based BCI, a number of stimuli are presented

to the user in a random order and the user draws his attention to a specific stimulus (usually by mentally counting its occurrences). After a number of repetitions of each stimulus, the system is able to predict on which stimulus the user was focusing on. The nature of the stimulus might be visual, auditory, tactile or combination of these. By altering his attention to different stimulus the user is able to perform different actions.

The most well-known P300-based multi-class BCI is the P300 speller proposed in 1988 by Farwell and Donchin Farwell and Donchin (1988). In the typical P300-speller paradigm the user stares at a screen where the characters are placed on a grid. As the characters are flushing in a random order, the user focuses on the character he/she wants to spell. Every time the attended character flashes, a P300 potential is generated. After a number of repetitions, the character that causes the stronger P300 peaks is classified by the system as the attended character. Implementations of the P300-speller have also been proposed using auditory instead of visual stimuli (e.g. Schreuder et al. (2011); Käthner et al. (2013)).

Apart from typing, a big variety of P300-based BCIs - targeted mainly for locked-in patients- has been proposed, such as controlling the mouse cursor (e.g. Citi et al. (2008)), controlling an internet browser (e.g. Mugler et al. (2010)), controlling a wheelchair (e.g. Rebsamen et al. (2007)), painting (e.g. Laar et al. (2013)), or controlling musical interfaces (e.g. Grierson (2008); Chew and Caspary (2011)).

Grierson (2008) presented a P300-based BCI where the user selects the midi-note number placed on a grid, in a similar way a user spells letters in the P300 speller. The maximum speed achieved among 5 subjects was one note every 7 seconds. Another P300 based BCI proposed Chew and Caspary (2011), integrates the idea of the P300 speller in a music 8x8 step sequencer. The notes of the sequencer are flashing in a random order, and the user selects them as he/she would select letter in the speller. At the same time the melody produced by the sequencer is played back. These last two proposed interfaces use only visual stimuli for controlling the musical interface. The following proposed P300-based BCMIs are can be controlled using either only the auditory modality or auditory and visual modality during the stimuli presentation.

We evaluate a proposed P300 BCMI using the Emotiv Epoc headset, and in another similar interface using the Enobio 8 wireless EEG headset.

4.8.1   Brain Sequencer

This application is designed for constructing arpeggios. The user looks at a 6x6 step sequencer on a computer screen. The buttons of the first column of the sequencer start flashing in a random order with Inter Onset Interval (IOI) 300 ms, while at the same time the note assigned to them sounds. The number of repetitions in the oddball paradigm might vary, depending on the desired accuracy.

Figure 4.24 shows a user performing with the interface. The tuning of the notes and the used timbre might also be set by the user. A recorded performance can be found online at https://youtu.be/PRGHeeO0WMk. In this performance the IOI was set to 300 ms and the available notes were these of a pentatonic scale (c, d, e, g, a). Any of the auditory or visual feedback modality can be turned-off. This was the first preliminary interface we implemented, and no proper evaluation was performed. We thought that interfaces in which the stimuli presentation is the musical outcome would be more interesting.

4.8.2   P300 harmonies

The interface described in this subsection is published at the joint Proceedings of the 40th International Computer Music Conference, ICMC 2014 and 11th Sound and Music Computing Conference (Vamvakousis and Ramirez (2014)).

The Interface

The interface consists of an arpeggio of six notes that is continuously being played back. The notes of the arpeggio sound in a random order. The arpeggio consists of 6 notes separated by an interval of 175ms. The notes of the arpeggio are controlled through 6 switches, where each switch has two possible states: up and down. When a switch is in the up-state the note produced by this switch is one tone or semitone -depending on the switch- higher than when in the down-state. By focusing on each of the notes of the arpeggio, the user may change -after 12 repetitions- the state of the corresponding switch. The state of each switch is shown on a screen (see figure 4.25). Each switch flashes when the corresponding note is heard. The user can either focus exclusively on distinguishing the desired sound or focus as well on the flashing of its corresponding switch. When all notes of the arpeggio have sounded 12 times, the background color of the screen

changes, indicating that the user can then focus on the next sound he desires to change.

In figure 4.25 are shown the notes assigned to each switch. When all switches are placed in the down-position, the re- sulting arpeggio consists of the notes G3 (sol in the 3rd octave), B3, D4, F#4, B4, D5, resulting in a G Major sev- enth chord, while when all switches are in the up-state, the arpeggio consists of the notes A3, C4, E4, G4, C5, E5, resulting in a A minor/minor seventh chord. Stereo spa- tialization is applied to the notes: the low pitch notes are placed to the left while as the pitch goes higher, the spa- tialization moves to the right. The interface has been tried so far with a sound of a harp.

By switching his attention to the notes of the arpeggio, the user can build a big variety of possible harmonies. The advantage of the proposed interface, when compared to previously proposed P300-based Musical Interfaces is that it can depend only on the auditory modality: the users changes the music, only by listening to it. Moreover, there is no time interval between the trials, resulting in a continuous musical outcome.

In the initial state all switches are placed down. Once the arpeggio starts being reproduced, every 72 notes (12 occurrences of each one of the 6 stim- uli), the background color of the screen changes, indicating that the user might then attend the next note he/she wishes to change. After about 1 second the voting classifier outputs the detected target stimulus, changing the state of the corresponding switch. As a result a different harmony is being produced by the arpeggio. This process, allows a continuous play- back of the musical outcome of the interface. The number of trials -that determines the duration of the performance- has to be determined at the beginning of the session.

Classification Process

The classification process is the same as the one described in 4.6.1. At the beginning of a training session one of the 6 notes of the arpeggio is played back to the user. After a small interval of 3 seconds, the stimuli are presented in a random order, under the constraint that at least one note interferes between two occurrences of the same note. The user is asked to mentally count the occurrences of the target-stimulus. A stimulus consists of the sound of the note, along with a blink on the screen of duration 100ms of the corresponding switch. The Inter-Stimulus- Interval (ISI) is set to 175ms. All stimulus are presented 12 times, until the next target stimulus

is presented to the user. This process is repeated 6 times -one for each stimulus-. As a result, the training data consist of 432 epochs, 72 of which are target epochs.

An epoch consists of the 14-channel recording of the time interval 250 to 750ms after the presentation of a stimu- lus. The signal is downsampled to 32Hz and band-pass filtered to 1-12Hz. Using the xDAWN (Rivet et al. (2009)) Spatial Filter Trainer in Openvibe, a 14 to 3 channels spatial filter is acquired. The 48 resulting values per epoch are then used to obtain a two-class Linear Discriminate Analysis Classifier (LDA) to distinguish target from non-target epochs. Once the spatial filter and the LDA classifier parameters are acquired, the use might start using the interface.

During the on-line session, the features per epoch, are being produced as in the training session. Then, for each stimulus a voting classifier computes the sum of the hyperplane distances -given by the LDA classifiers-, and outputs as the attended stimulus the one with the lowest sum.

Evaluation and Results

The interface was evaluated with 4 subjects (3 male). After training the system -as described in paragraph 2.2.1- they were asked to move all switches up, starting from the leftmost one and moving to the one in the right. The average age of all subjects was 35 years. The only female subject performed the task in an exhibition setting, using loudspeakers for sound generation achieving 100% accu racy 5 . The 3 remaining subjects were asked to perform the same task in an office environment, using in-ear head- phones. The accuracy was 6/6, 4/6 and 5/6. All subjects used both the visual and auditory modality of the interface to control the interface. The average ITR among all 4 subjects was 7.37 bits/min, while in the case of the 2 subjects that performed with 100% accuracy the achieved ITR was 12.31 bits/min.

Discussion

The novelty of the proposed interface lies in the fact that the user voluntarily interacts with the music while listening to it. The limitation that a P300-based auditory BCMI introduces is that the stimuli should be presented in a random order. Even with this limitation though, interesting musical interfaces can be designed. Such interfaces could be useful for some cases of locked-in patients. In the proposed BCI, the stimuli presentation of a trial starts before presenting the outcome of the preceding. Due to this

	Harp/-90	Harpichord/-45	Harp/o	Harpichord/45	Harp/90
F	a3	c4	f4	a4	c5
C	g3	c4	e4	g4	c5
G	g3	b3	d4	g4	b4
Dm	a3	d4	f4	a4	d5
Am	a3	c4	e4	a4	c5

Table 4.11: The notes corresponding to each chord. The name of the note is followed by its octave. For example a3 corresponds to the note a at the 3rd octave. The musical instrument used and the stereo panning of each stimulus are also shown.

fact, the Information Transfer Rate of the system increases when compared to a system where a time interval is introduced between the trials. The preliminary evaluation presented here, did not study the accuracy of the system when just the auditory modality is used in stimuli presentation. We studied this in a different version of the interface presented in the following section.

4.8.3   BrainLoops

The interface described in this subsection is published at the Proceedings of the 1st International BCMI Workshop (Vamvakousis and Ramirez (2015)).

The Interface

The interface consists of a dynamically changing 5-class P300 oddball paradigm. Two different scenarios were implemented: the audiovisual and the auditory. Table 4.8.3 summarizes the timbre of each stimulus (which is the musical instrument that generates its sound), along with its stereo panning and the pitch. The pitch of each stimulus varies, depending on which is the currently selected chord.

During a trial all stimuli appear in a random order, 20 times each. The Inter Onset Interval (IOI) (time interval between the onset of two consecutive stimuli) was set to 220ms and the duration of each stimulus to 80ms. At least one stimulus interferes between two occurrences of the same stimulus. Each stimulus is mapped to a different chord. In an order from left to right in the stereo panning position, the stimulus are mapped to the following chords: F major, C major , G major , D minor and A minor. If during a trial the user wants to select one of the chords, he/she has to switch his

attention to the corresponding stimulus and mentally count its occurrences. At the end of each trial the system detects which was the attended stimulus and tunes all stimuli to the corresponding chord. There is no time interval between two trials. Instead, after switching to the selected chord, each stimulus appears three time, with an order from left to right. This helps the user to spot the next stimulus he/she wishes to attend. In front of the user there is a computer screen showing the available chords to be selected. The chords are placed in the same order as the corresponding stimuli (see figure 4.26). The currently selected chord is marked in red. Just in the case of the audiovisual scenario, the chord names are flashing when the corresponding stimulus sounds. In the stimuli design there are 3 discriminating cues: timbre, pitch and stereo spatialization. From left to right, the sound of a harp alternates with a sound of a harpsichord in a way that two neighboring sounds do not share the same timbre. Both of the selected instruments are stringed and have similar timbre. This was preferred to preserve the musicality of the interface.

Classification Process

Before using the interface in the online session, a training session is performed in order to acquire data to obtain a two class LDA classifier and a spatial filter, as explained in 4.6.1. During the training session an arrow is pointing at the stimulus the user should attend. The training session consists of 10 trials, in which each stimulus is set as a target twice.

Evaluation and Results

The Enobio 8 wireless electrophysiology sensor systemwas used for recording the brain signals. The proposed interface was evaluated on 8 healthy subjects (5 male, mean age 29 years, standard deviation 5.46). All subjects gave oral consent to participate in the study. The Enobio 8 wireless electrophysiology sensor systemwas used for recording the brain signals.

On each user it was first evaluated the audiovisual and then the auditory scenario. For each scenario a 10-trials training session was followed by a 15-trials online session. In order to evaluate the accuracy of the system, the users were asked to select all the chords with order from left to right as shown in figure 1. As the training session consisted of 10 trials, each stimulus was set as target in 2 trials in that case, while it was set as target 3 times in the online session. An arrow was always pointing to the stimulus to be attended. The first two repetitions of each stimulus were not taken

	Audiovisual		Auditory	
User	10-fold Cross Validation	Accuracy out of 15	10-fold Cross Validation	Accuracy out of 15
M32	65.9%	8	63.1%	12
M27	71%	14	64.7%	5
M34	66.9%	14	62.7%	7
M38	65.3%	8	59.8%	5
F20	70.5%	12	58.3%	8
F27a	66.2%	11	61.1%	7
M28	77.5%	15	60.9%	3
F27b	70.10%	14	57.3%	6
AVG	69.18%	12 (80%)	60.99%	6.63 (44%)

Table 4.12: The 10-fold Cross Validation and selection accuracy for all users in both the audiovisual and auditory scenario.

into account during the classification process. The reason for that is that presumably it required some time for the users to spot the desired stimulus. Table 4.8.3 summarizes the selection accuracy and 10-cross fold validation of all users. The average selection accuracy for the audiovisual scenario is 80%, while for the auditory scenario it is 44%.

Discussion

In this study we implemented and evaluated an P300-based BCMI and evaluated it in two scenarios: the audiovisual and the auditory. In our paradigm the user is able to change the harmony of the sound stimuli by switching his attention to each one of them. This is a special case of an auditory oddball paradigm, where the properties of the stimuli change according to the selections of the user. This probably makes the oddball task even more difficult. In the audiovisual scenario, where the users also make use of the visual cue, the selection accuracy is 80%. When the visual cue is removed, the selection accuracy falls to 44%. In the audiovisual scenario all users reported that they counted all 20 occurrences of the attended stimulus in every trial. On the contrary in the case of the auditory scenario several subject reported that in some of the trials they were not able to count all occurrences of the attended stimulus. Subject F27a reported that she was confusing the 3rd with the 4th stimulus (as they are placed in the stereo panning from left to right). The same was reported by subject M28. Sub-

ject F27b reported that was counting about 18 occurrences instead of 20 in all the trials. Subject M27 could not count correctly the 3rd stimulus in one of the trials. The rest of the subjects reported that they were able to count 20 trials of the target stimulus in all trials. These subjects happen to have musical training. The auditory scenario was more difficult for all subjects, apart from the subjects with musical training (M32, M38, M34 and F20). The difficulty of the subjects with no musical training to spot the stimulus in the auditory sce- nario is also reflected in their selection accuracy. Musicians average 50% while non-musicians 33.3%. Subject M32 was the only subject that had previous experience with auditory P300-based BCIs. He was the only one that achieved a high selection accuracy. This indicates that training might be crucial in such interfaces.

The lower performance of auditory P300-based BCIs compared to the visual ones is known by previous studies. In the current study though we observe an even higher difference. This could be explained by the fact that the sound stimulus are dynamically changing according to the selections of the user. Another explanation is that the musicality of the interface was taken into account in stimuli design. The musical interval between to neighbor stimulus was a 3rd or 4th, while all stimuli are always harmonic with each other. Dissonant notes might have been easier to distinguish. Another parameter is the selected timbre of the stimuli. Instead of using a different musical instrument for each stimulus, we only selected two different instruments with similar timbre in order to preserve the musicality of the interface. A way to improve the system's accuracy could by increasing the number of repetitions of the oddball paradigm. This would make the interaction slower. Twenty-two seconds per selection is already a big time interval. Given the results of the current study, implementing an auditory P300-based BCMI is hardly feasible.

4.9 Conclusions

In this chapter we investigated different ways of implementing Brain Computer Musical Interfaces. We described a simple application in which the user controls the contour of a melody by performing imagery feet movements. Then we presented an emotion estimation method based on EEG data using Emotiv Epoc headset. Our results indicate that EEG data obtained with Emotiv Epoc device contain sufficient information to distinguish these emotional states, and that machine learning techniques are capable of learning the patterns that distinguish those states.

We applied this emotion-estimation technique in a hybrid brain-gaze controlled music performance, in which the performed controlled the melody with the eyes, while his detected emotional state was driving an arpeggio generation system.

Then we investigated auditory Event Relate Potential interfaces. In particular we evaluated the performance of low cost Emotiv Epoc headset in capturing auditory evoked potentials. Results indicate that Emotiv Epoc can be used in auditory ERP-based BCIs. We then explored how different auditory cues (timbre, pitch and spatialization), and combination of those, affect the accuracy of auditory ERP-based BCIs. Results indicate that when combining all auditory cues in the stimuli design, the performance is maximized.

Afterwards we proposed two auditory ERP-based interfaces that allow the control of the harmony of the auditory stimuli presentation. Results so that when only the auditory modality is used, the accuracy in this case is low even when using a high quality EEG device.
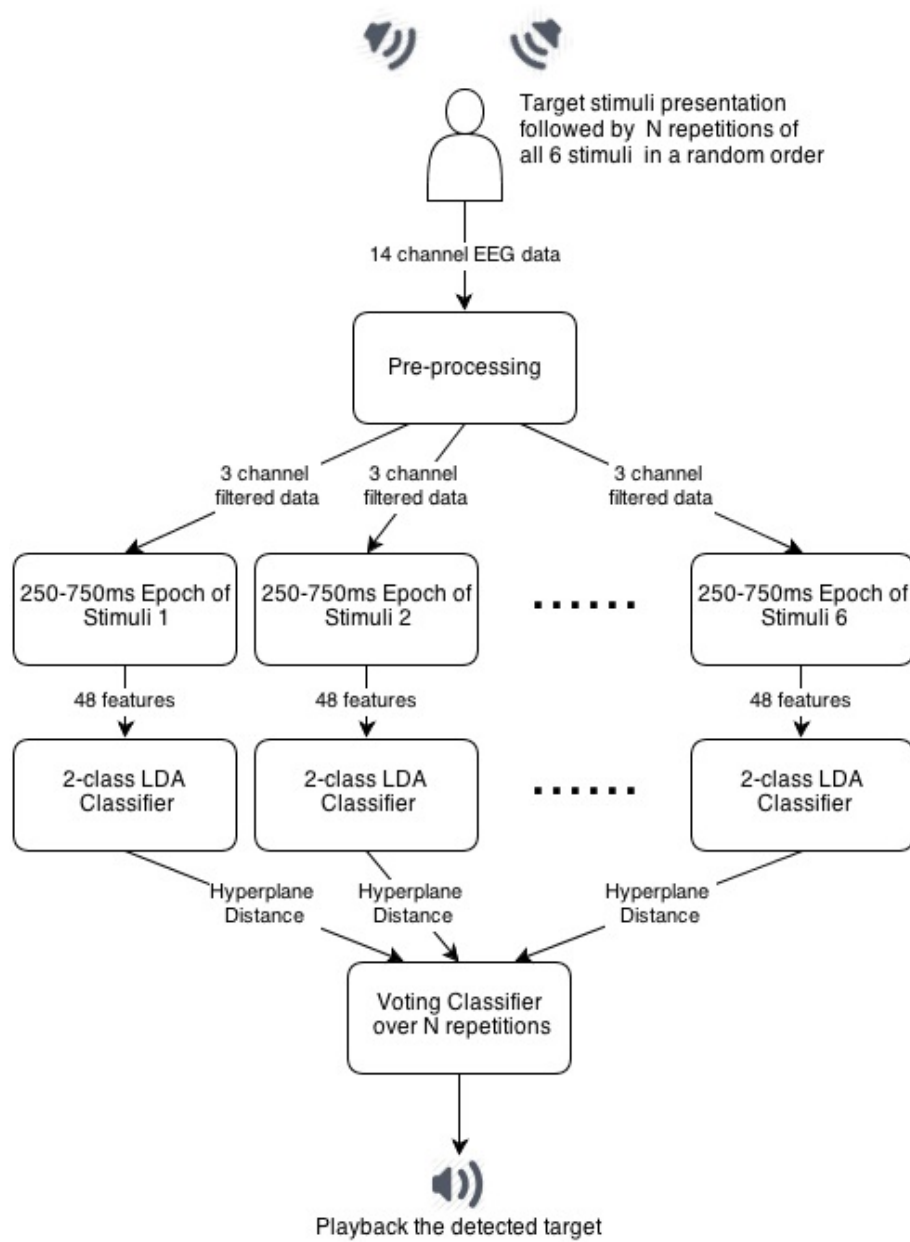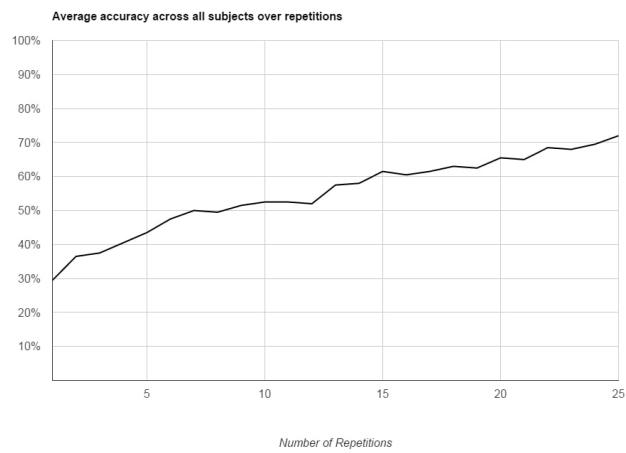
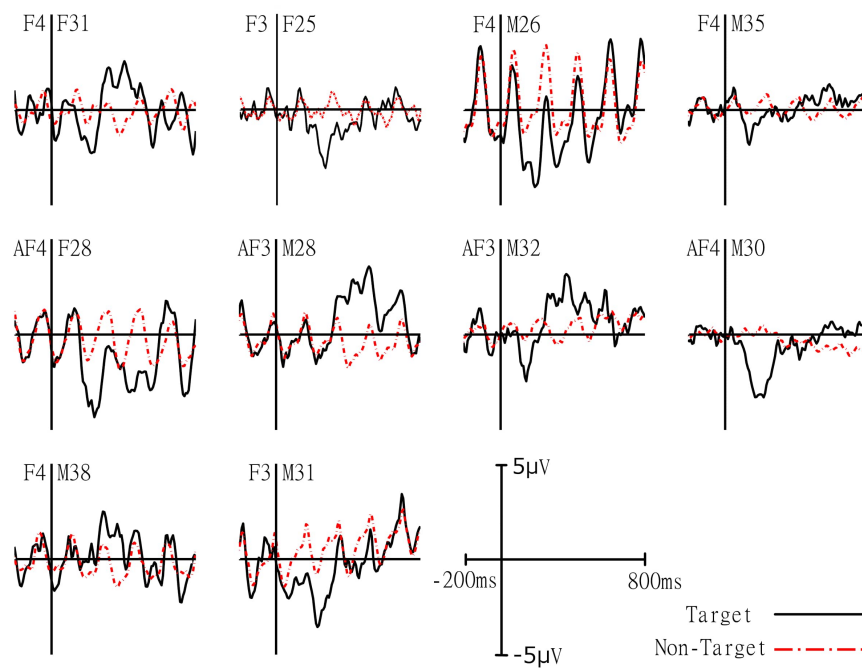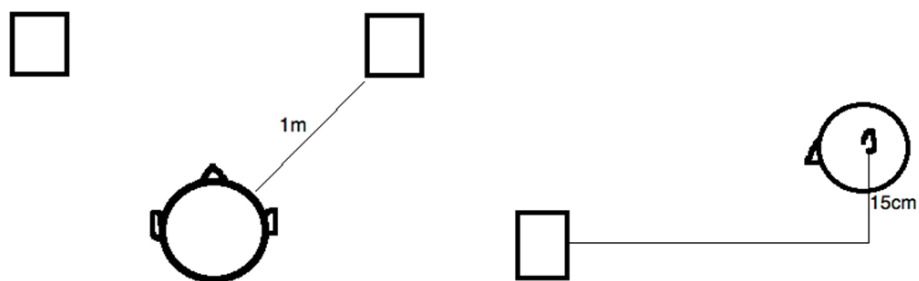**Figure 4.12:** Online classification process.

Figure 4.13: ERS observed in the motor cortex of a subject during real feet movement

Figure 4.14: Averaged target and non-target waveforms for all users. Only the location with maximum target and non-target difference is shown for each subject.

Figure 4.15: Experiment setup: for all experiments two loudspeakers were used to spatialize the stimuli
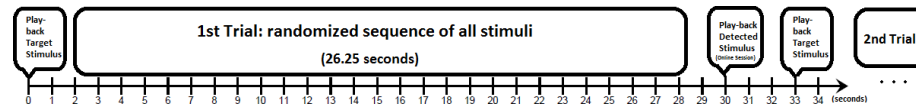
Figure 4.16: A session of the 175 ms ISI conditions. Each session consisted of 10 trials. . Before each trial, a random stimulus was played back as the target stimulus. In the case of 175ms ISI conditions a trial consisted of 25 repetitions of all stimuli in a random order and lasted for 26.25 seconds. In the case of 300ms (15 repetitions) and 150ms (25 repetitions) ISI conditions each trial lasted 27 and 22.5 seconds, respectively. In the on-line sessions, the detected target stimulus was played-back after each trial.
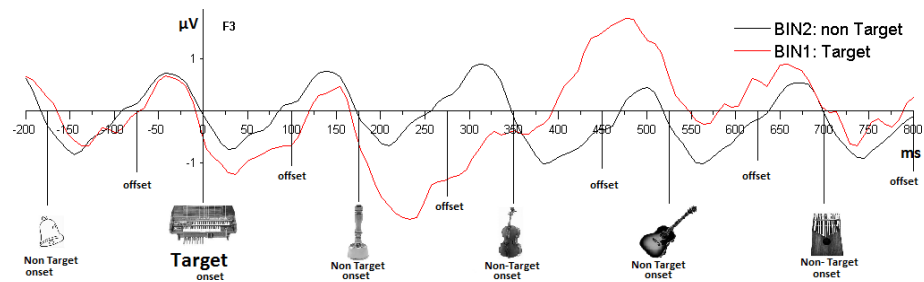
Figure 4.17: Stimuli presentation of a repetition for the TimPiSp175 condition. Additionally the averaged ERP response measured in the F3 channel of all users for the TimPiSp175 condition is shown. The red line corresponds to the target epochs and the black line corresponds to the non-target epochs. The ERP responses follow the periodicity of the stimuli presentation.
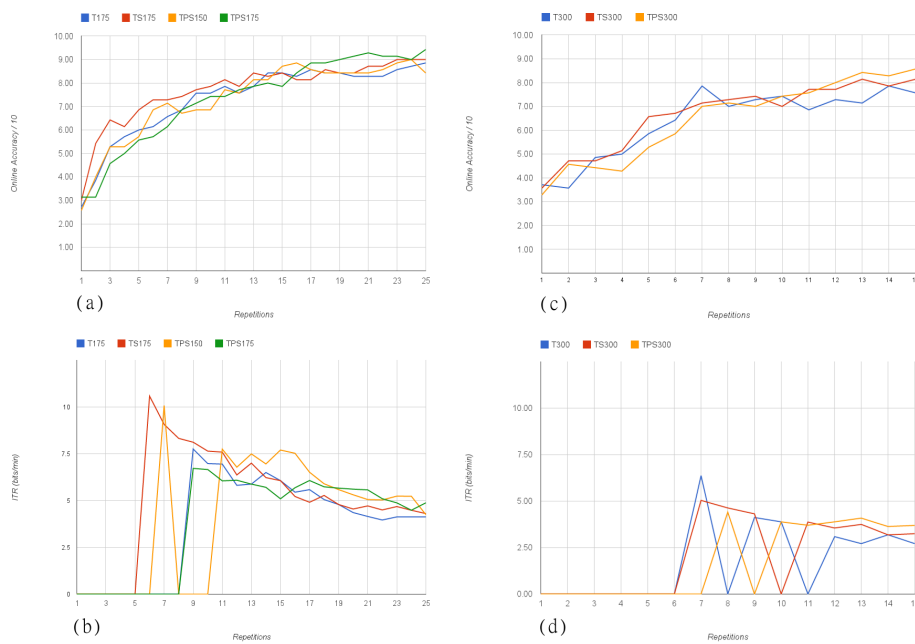
Figure 4.18: (a,b) Averaged on-line performance and ITR of all subjects for the 175 and 150ms conditions for different number of repetitions. (c,d) Averaged on-line performance and ITR of all subjects for the 300ms conditions for different number of repetitions. The ITR is considered to be zero, if the average accuracy is less than 70%.

m46 F4   m30 F8   m36 F4   f28 F4

m28 AF4   f58 F3   F57 F7
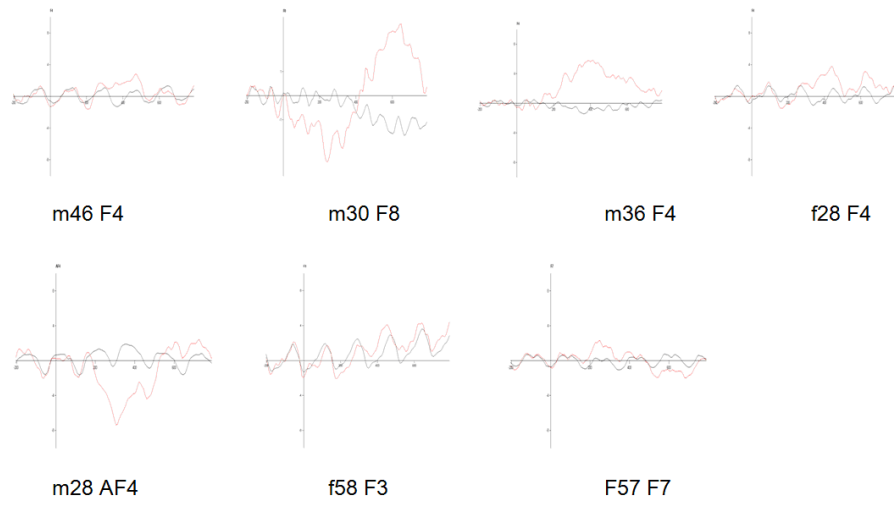
Figure 4.19: 175ms ISI grand average.
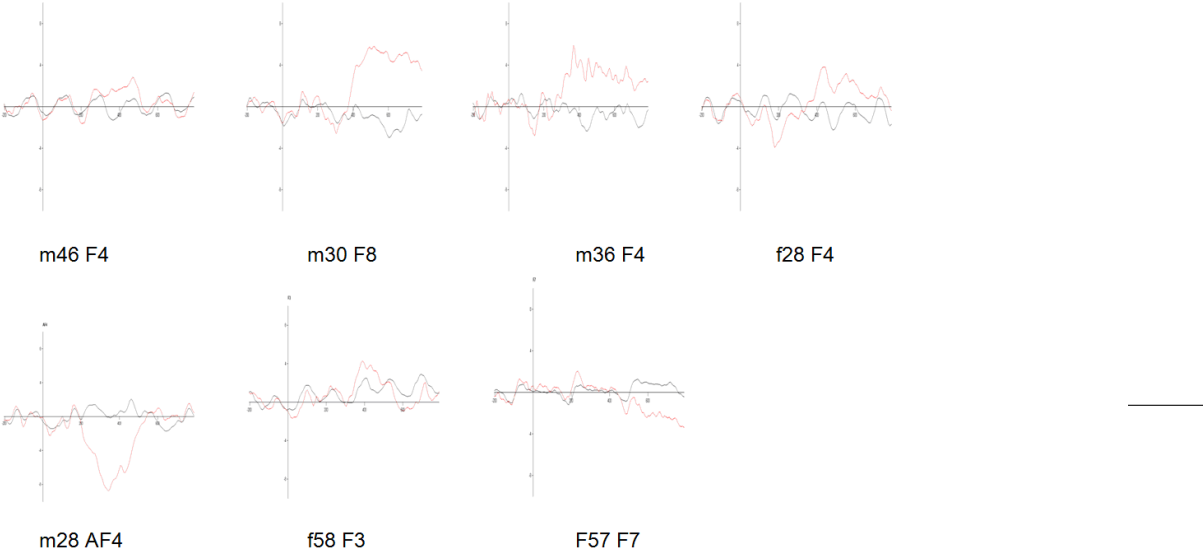
Figure 4.20: 100ms ISI grand average.
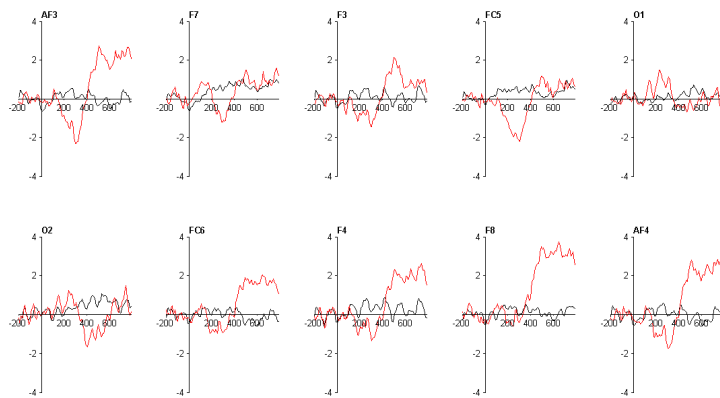
Figure 4.21: 300msTPS all subjects 10 electrodes average.
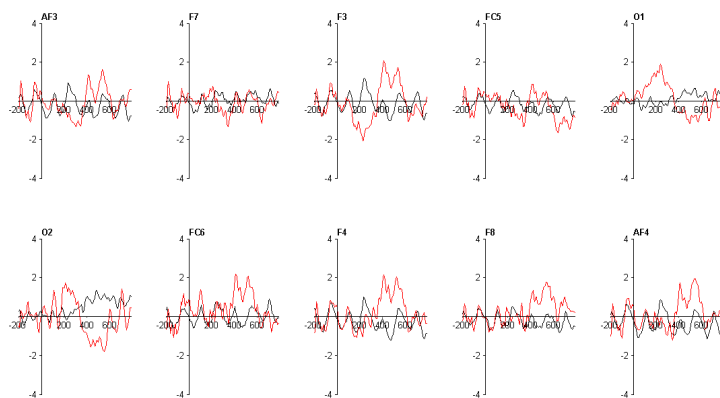
Figure 4.22: 300msTS all subjects 10 electrodes average
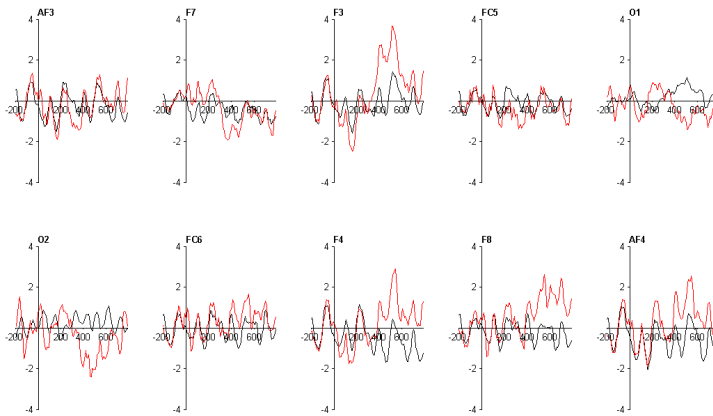
Figure 4.23: 300msT all subjects 10 electrodes average
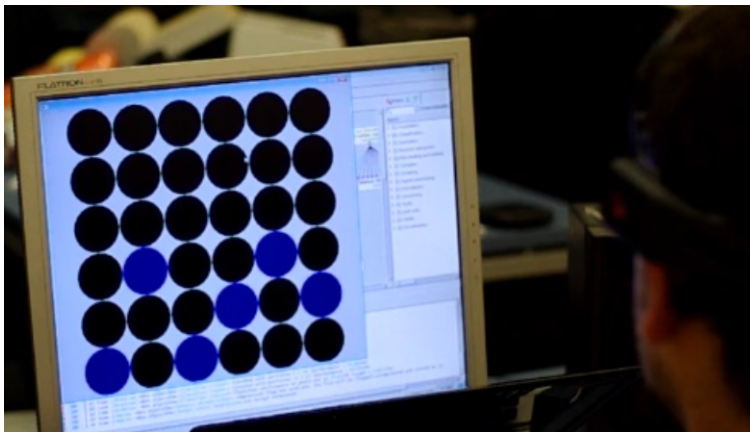
Figure 4.24: Brain Sequencer: Constructing Arpeggios using auditory or auditory+visual stimuli as an input in a P300 oddball task. The user inputs the notes one by one. First the note of the first column is selected, followed by the note of the second column and so on.

Figure 4.25: From each switch the user can select between two possible notes. The selected note of each switch is highlighted in blue color. When the program starts, all switches are placed down.
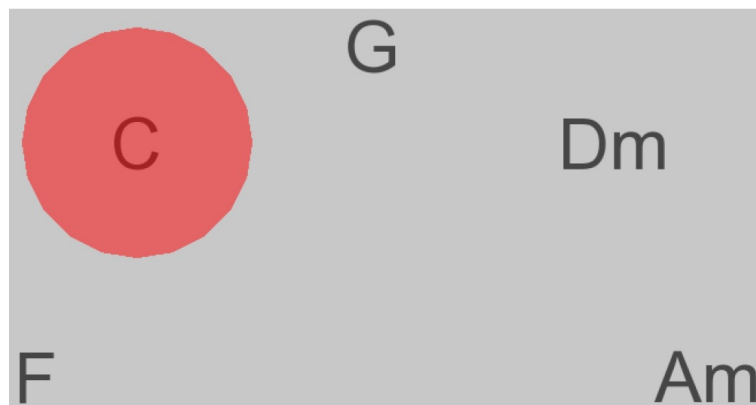
Figure 4.26: The visual feedback. In the case of the audiovisual scenario the chord names flash when the corresponding stimulus sounds. The red circle indicates which is the currently selected chord.

Conclusions

5.1 Introduction

In this dissertation we have investigated the feasibility of creating digital musical interfaces designed for people with different types of physical disabilities.

In chapter 2 we studied three different cases of people with limited upper limb sensorimotor functions. We have shown that using the arduino microcontroller with low cost sensors and materials, it is possible to construct Digital Musical Instruments that might allow people with physical disabilities to perform music. There are indications that this has a positive impact in their quality of life. In particular in the case of the study presented in section 2.2, while the user was not able to play any music instrument before providing him with the implemented prototype, he now sees his future as a professional music composer, while he has already composed music for a theatrical play. He states that probably without the constructed midi controller he would not be able to improvise and generate ideas for his compositions. In the case of the study presented in section 2.4, although the user does not seem to improve much in his playing abilities, the constructed prototype seems to have a positive impact therapeutically in his life. Finally, in the case of the study presented in chapter 2.3 the final constructed prototype had some elements that should be kept, but some basic design choices should be reconsidered in order to make it usable for the subject.

In chapter 3 we presented the EyeHarp gaze-controlled music interface. We studied the temporal accuracy in playing simple music exercises. We proposed and evaluated a novel fixation-detection algorithm that according to

the results provides better temporal accuracy than other commonly used algorithms. Then we evaluated the EyeHarp as a Digital Musical Instrument from the perspective of the audience and of the performer. Results indicate that the EyeHarp is a DMI with steep learning curve that allows expressive performances.

In chapter 4, we examined the possibility of brain-controlled music interfaces. Initially we experimented with a sensorimotor rhythm-based music interface. We then evaluated the potential of estimating emotional states out of the EEG signal, using the Emotiv Epoch device. Our results indicate that it is possible to detect basic emotion by applying the methodology we propose. As a proof of concept, he organized a concert in which within the EyeHarp interface the performer was the solo part, while the detected emotional states were driving an arpeggio generation system. We then evaluated the performance of Emotiv headset in auditory Event-Related Potential based brain-computer interfaces. The results indicate that Emotiv Epoc can be used successfully in such interfaces. Additionally, we investigated how timbre, pitch and spatialization auditory cues in the stimuli design affect the performance of such interfaces. Our results propose that by combining all auditory cues, the performance of such interfaces is maximized. Finally, we proposed three different auditory ERP-based brain-computer music interfaces. We evaluated their accuracy and Information Transfer Rate in two scenarios: (i) when stimuli consisted of only auditory cues and (ii) when stimuli were presented from both the auditory and the visual channel. Our results indicate that in the audio-visual scenario, such an interface might have a robust performance. Nevertheless, our results indicate that the performance of an ERP-based BCMI using only the auditory channel is low for implementing usable BCMIs.

5.2 Summary of Contributions

In this dissertation we contributed in three distinct fields of research that are related to interfaces for people with physical disabilities:

Limb-Controlled

- We confirmed that using low cost materials, digital musical instruments designed for people with upper-limb paresis can be constructed and serve them for music expression and composition.

- Publicly available instructions of how to implement the constructed prototypes.

Gaze-Controlled

- We implemented and evaluated the EyeHarp, an open-source gaze-controlled music interface. Results indicate that it offers a steep learning curve and it allows expressive live performances.

- A new fixation-detection algorithm was implemented and evaluated. Results indicate that it provides improved temporal accuracy when compared to commonly used fixation-detection algorithms.

- Different gaze-selection techniques for improving the spatial accuracy in selecting small targets were implemented and compared with each other.

Brain-Controlled

- Our results indicate that it is possible to estimate emotional states using the low cost Emotiv Epoch EEG device.

- Our results indicate that the performance auditory Event-Related Potential Brain Computer Interfaces is maximized when the used stimuli vary in timbre, pitch and spatialization.

- Our results indicate that Emotiv Epoch EEG device can be effectively used in auditory ERP-based BCIs.

- All recorded EEG data from our experiments are publicly available. EEG recording available at: [1]

- Auditory and audiovisual Event-Related Potential-based interfaces and a hybrid Brain-Gaze interface were proposed as possible music interfaces for people in total locked-in state.

---

[1] https://drive.google.com/open?id=0B6BzfrftsEXWTHlHMC1oUHE4RU0

Bibliography

Each reference indicates the pages where it appears.

Daniel Arfib, Jean-Michel Couturier, and Loïc Kessous. Expressiveness and Digital Musical Instrument Design. *Journal of New Music Research*, 34(1):125–136, 2005. ISSN 0929-8215. doi: 10.1080/09298210500124273. URL http://www.tandfonline.com/doi/abs/10.1080/09298210500124273?src=recsys{#}.VQHeLPmsWWs. 58, 64, 68

Nicholas A Badcock, Petroula Mousikou, Yatin Mahajan, Peter de Lissa, Johnson Thie, and Genevieve McArthur. Validation of the Emotiv EPOC(®) EEG gaming system for measuring research quality auditory ERPs. *PeerJ*, 1:e38, jan 2013. ISSN 2167-8359. doi: 10.7717/peerj.38. URL https://peerj.com/articles/38. 94, 107

Marc Bangert and Eckart O Altenmüller. Mapping perception to action in piano practice: a longitudinal DC-EEG study. *BMC neuroscience*, 26(4):1–14, 2003. ISSN 1471-2202. doi: 10.1186/1471-2202-4-26. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=270043{&}tool=pmcentrez{&}rendertype=abstract. 2

J Barbosa and F Calegario. Considering audience's view towards an evaluation methodology for digital musical instruments. *NIME'12 Proceedings . . .*, 2012. URL http://batebit.cc/wp-content/uploads/2014/07/174{_}final{_}manuscript.pdf. 58, 59

Jessica D Bayliss, Samuel A Inverso, and Aleksey Tentler. Changing the P300 brain computer interface. *Cyberpsychology & behavior : the impact of the Internet, multimedia and virtual reality on behavior and society,*

7(6):694–704, dec 2004. ISSN 1094-9313. doi: 10.1089/cpb.2004.7.694.
URL http://www.ncbi.nlm.nih.gov/pubmed/15687805. 91

Mireille Besson and Daniele Schön. Comparison Between Language and
Music. *The Cognitive Neuroscience of Music*, pages 232–258, 2012. ISSN
0077-8923. doi: 10.1093/acprof:oso/9780198525202.003.0018. 2

Margrit Betke, James Gips, and Peter Fleming. The Camera Mouse:
Visual tracking of body features to provide computer access for peo-
ple with severe disabilities. *IEEE Transactions on Neural Systems and
Rehabilitation Engineering*, 10(1):1–10, 2002. ISSN 15344320. doi:
10.1109/TNSRE.2002.1021581. 30

Soumitra Bhat. TouchTone An Electronic Musical Instrument for Children
with Hemiplegic Cerebral Palsy. *TEI '10: Proceedings of the fourth in-
ternational conference on Tangible, embedded, and embodied interaction*,
pages 305–306, 2010. doi: 10.1145/1709886.1709955. 10

Niels Birbaumer, Andrea K??bler, Nimr Ghanayim, Thilo Hinterberger,
Jouri Perelmouter, Jochen Kaiser, Iver Iversen, Boris Kotchoubey, Nicola
Neumann, and Herta Flor. The thought translation device (TTD) for
completely paralyzed patients. *IEEE Transactions on Rehabilitation En-
gineering*, 8(2):190–193, 2000. ISSN 10636528. doi: 10.1109/86.847812.
73

Yvonne Blokland. Imagined music as a paradigm for brain-computer inter-
faces. Master's thesis, Utrecht University, 2009. 73, 76

Danny Oude Bos. EEG-based Emotion Recognition: The Influence of Visual
and Auditory Stimuli. *(?)*, pages 1–17, 2006. ISSN 15582531. doi: 10.
1109/TBME.2010.2048568. 81

A S Chan, Y C Ho, and M C Cheung. Music training improves verbal
memory. *Nature*, 396(6707):128, 1998. ISSN 0028-0836. doi: 10.1038/
24075. 2

Guillaume Chanel, Julien Kronegg, Didier Grandjean, and Thierry Pun. Pe-
ripheral Physiological Signals Emotion Assessment : Arousal Evaluation
Using EEG ' s and Peripheral Physiological Signals Computing Science
Center , University of Geneva. *Mulimedia Content Representation, Clas-
sification and Security.*, 2006. 81

Yee Chieh Chew and Eric Caspary. MusEEGk: a brain com-
puter musical interface. *Proceedings of the 2011 annual confer-
ence extended abstracts on Human factors in computing systems
- CHI EA '11*, pages 1417–1422, 2011. ISSN 9781450302685.
doi: 10.1145/1979742.1979784.   URL   http://dl.acm.org/

`citation.cfm?id=1979784$\delimiter"026E30F$nhttp://doi.`
`acm.org/10.1145/1979742.1979784$\delimiter"026E30F$nhttp:`
`//portal.acm.org/citation.cfm?doid=1979742.1979784`. 77, 117

A Choppin. Eeg-based human interface for disabled individuals: Emotion expression with neural networks. *Unpublished masters thesis*, 2000. 81

Luca Citi, Riccardo Poli, Caterina Cinel, and Francisco Sepulveda. P300-based BCI mouse with genetically-optimized analogue control. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 16(1): 51–61, 2008. ISSN 15344320. doi: 10.1109/TNSRE.2007.913184. 117

Kerry L Coburn and Marco A Moreno. Facts and artifacts in brain electrical activity mapping. *Brain Topography*, 1(1):37–45, 1988. 84

Don D Coffman. Music and quality of life in older adults. *Psychomusicology: A Journal of Research in Music Cognition*, 18(1-2):76, 2002. 2

H Collewijn, C J Erkelens, and R M Steinman. Binocular coordination of human horizontal saccadic eye movements. *The Journal of physiology*, 404:157–82, 1988a. ISSN 0022-3751. doi: 10/ VL-404. URL `http://www.pubmedcentral.nih.gov/articlerender.` `fcgi?artid=1190820{&}tool=pmcentrez{&}rendertype=abstract`. 52

H Collewijn, CASPER J Erkelens, and RM Steinman. Binocular coordination of human vertical saccadic eye movements. *Journal of Physiology*, 404(1):183–197, 1988b. 52

Shirley M Coyle, Toms E Ward, and Charles M Markham. Braincomputer interface using a simplified functional near-infrared spectroscopy system. *Journal of Neural Engineering*, 4(3):219, 2007. URL `http://stacks.` `iop.org/1741-2552/4/i=3/a=007`. 71

Nello Cristianini and John Shawe-Taylor. *An introduction to support vector machines and other kernel-based learning methods*. Cambridge university press, 2000. 86

Maarten De Vos, Katharina Gandras, and Stefan Debener. Towards a truly mobile auditory brain-computer interface: exploring the P300 to take away. *International journal of psychophysiology : official journal of the International Organization of Psychophysiology*, 91(1):46–53, jan 2014. ISSN 1872-7697. doi: 10.1016/j.ijpsycho.2013.08.010. URL `http://www.` `sciencedirect.com/science/article/pii/S0167876013002468`. 103, 105

Arnaud Delorme and Scott Makeig. EEGLAB: an open sorce toolbox for analysis of single-trail EEG dynamics including independent component anlaysis. *Journal of Neuroscience Methods*, 134:9–21, 2004. ISSN 0165-

0270. doi: 10.1016/j.jneumeth.2003.10.009. 111

Leon Y. Deouell, Ariel Parnes, Natasha Pickard, and Robert T. Knight. Spatial location is accurately tracked by human auditory sensory memory: Evidence from the mismatch negativity. *European Journal of Neuroscience*, 24(5):1488–1494, 2006. ISSN 0953816X. doi: 10.1111/j. 1460-9568.2006.05025.x. 92

Dimitri Diakopoulos and A Kapur. HIDUINO: A firmware for building driverless USB-MIDI devices using the Arduino microcontroller. *Proceedings of the International Conference on New Interfaces for Musical Expression*, (June):405–408, 2011. ISSN 2220-4806. URL http://dimitridiakopoulos.com/pubs/2011{_}nime{_}hiduino.pdf. 12

Matthieu Duvinage, Thierry Castermans, Mathieu Petieau, Thomas Hoellinger, Guy Cheron, and Thierry Dutoit. Performance of the Emotiv Epoc headset for P300-based applications. *Biomedical engineering online*, 12(1):56, jan 2013. ISSN 1475-925X. doi: 10.1186/1475-925X-12-56. URL http://www.biomedical-engineering-online.com/content/12/1/56. 94, 105

Adrian E.D. and Matthews B.H.C. The berger rhythm: Potential changes from the occipital lopes in man. *Brain*, 1934. 74

Gnter Edlinger, Cristiano Rizzo, and Christoph Guger. Brain computer interface. In Rdiger Kramme, Klaus-Peter Hoffmann, and Robert S. Pozos, editors, *Springer Handbook of Medical Technology*, pages 1003–1017. Springer Berlin Heidelberg, 2012. ISBN 978-3-540-74658-4. URL http://dx.doi.org/10.1007/978-3-540-74658-4_52. 10.1007/978-3-540-74658-4_52. 71

T Elbert, C Pantev, C Wienbruch, B Rockstroh, and E Taub. Increased cortical representation of the fingers of the left hand in string players. *Science (New York, N.Y.)*, 270(5234):305–307, 1995. ISSN 0036-8075. doi: 10.1126/science.270.5234.305. 2

L A Farwell and E Donchin. Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalography and clinical neurophysiology*, 70(6):510–23, dec 1988. ISSN 0013-4694. URL http://www.ncbi.nlm.nih.gov/pubmed/2461285. 91, 117

Mehrdad Fatourechi, Ali Bashashati, Rabab K Ward, and Gary E Birch. Emg and eog artifacts in brain computer interface systems: A survey. *Clinical neurophysiology*, 118(3):480–494, 2007. 84

A Furdea, S Halder, D J Krusienski, D Bross, F Nijboer, N Birbaumer,

and A Kübler. An auditory oddball (P300) spelling system for brain-computer interfaces. *Psychophysiology*, 46(3):617–25, may 2009. ISSN 1540-5958. doi: 10.1111/j.1469-8986.2008.00783.x. URL http://www.ncbi.nlm.nih.gov/pubmed/19170946. 91, 102, 103

Christian Gaser and Gottfried Schlaug. Brain structures differ between musicians and non-musicians. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 23(27):9240–9245, 2003. ISSN 1529-2401. doi: 23/27/9240[pii]. URL http://www.ncbi.nlm.nih.gov/pubmed/14534258. 2

Mick Grierson. Composing With Brainwaves : Minimal Trial P300 Recognition As an Indication of Subjective Preference for the Control of a Musical. *International Computer Music Conference*, 2008. 77, 117

S. Halder, M. Rea, R. Andreoni, F. Nijboer, E. M. Hammer, S. C. Kleih, N. Birbaumer, and A. Kübler. An auditory oddball brain-computer interface for binary choices. *Clinical Neurophysiology*, 121(4):516–523, 2010. ISSN 13882457. doi: 10.1016/j.clinph.2009.11.087. URL http://dx.doi.org/10.1016/j.clinph.2009.11.087. 92, 103, 105

Brahim Hamadicharef, Mufeng Xu, and Sheel Aditya. Brain-Computer Interface (BCI) based musical composition. *Proceedings - 2010 International Conference on Cyberworlds, CW 2010*, pages 282–286, 2010. doi: 10.1109/CW.2010.32. 77

John Paulin Hansen, Anders Sewerin Johansen, Dan Witzner Hansen, Kenji Itoh, and Satoru Mashino. Command Without a Click : Dwell Time Typing by Mouse and Gaze Selections. *Proceedings of Human-Computer Interaction INTERACT03*, pages 121–128, 2003. URL http://www.idemployee.id.tue.nl/g.w.m.rauterberg/conferences/interact2003/INTERACT2003-p121.pdf. 65

Eddie Harmon-Jones. Clarifying the emotive functions of asymmetrical frontal cortical activity. *Psychophysiology*, 40(6):838–848, 2003. 84

Terrence Hays and Victor Minichiello. The contribution of music to quality of life in older people: an Australian qualitative study. *Ageing and Society*, 25(2):261–278, 2005. ISSN 0144-686X. doi: 10.1017/S0144686X04002946. 68

Henrik Hegner and Tomra Skovsgaard. Estimating acceptable noise-levels on gaze and mouse selection by zooming. In *2008 Annual IEEE Student Paper Conference, AISPC*, pages 1–4. IEEE, feb 2008. ISBN 978-1-4244-2156-5. doi: 10.1109/AISPC.2008.4460544. URL http://ieeexplore.ieee.org.sare.upf.edu/articleDetails.jsp?arnumber=4460544. 42

T. Hinterberger and G. Baier. Parametric Orchestral Sonification of EEG in Real Time. *IEEE Multimedia*, 12(2):70–79, apr 2005. ISSN 1070-986X. doi: 10.1109/MMUL.2005.36. URL http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1423936. 74

Thilo Hinterberger, Nicholas Jeremy Hill, and Niels-Peter Birbaumer. An Auditory Brain-Computer Communication Device. *IEEE International Workshop on Biomedical Circuits and Systems, 2004*, pages S3/6/15—-18, 2004a. doi: 10.1109/BIOCAS.2004.1454171. URL http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1454171. 73, 78

Thilo Hinterberger, Nicola Neumann, Mirko Pham, Andrea Kbler, Anke Grether, Nadine Hofmayer, Barbara Wilhelm, Herta Flor, and Niels Birbaumer. A multimodal brain-based feedback and communication system. *Experimental Brain Research*, 154:521–526, 2004b. ISSN 0014-4819. URL http://dx.doi.org/10.1007/s00221-003-1690-3. 10.1007/s00221-003-1690-3. 75

A. S. Ho, Y. C., Cheung, M. C., Chan. Music training improves verbal but not visual memory: Cross-sectional and longitudinal explorations in children. *Neuropsychology*, 17(3):439–450, 2003. ISSN 0894-4105. doi: 10.1037/0894-4105.17.3.439. 2

Leigh R. Hochberg, Daniel Bacher, Beata Jarosiewicz, Nicolas Y. Masse, John D. Simeral, Joern Vogel, Sami Haddadin, Jie Liu, Sydney S. Cash, Patrick van der Smagt, and John P. Donoghue. Reach and grasp by people with tetraplegia using a neurally controlled robotic arm. *Nature*, 485(7398):372–375, 2012. ISSN 0028-0836. doi: 10.1038/nature11076. URL http://dx.doi.org/10.1038/nature11076. 73

Johannes Höhne, Martijn Schreuder, Benjamin Blankertz, and Michael Tangermann. A Novel 9-Class Auditory ERP Paradigm Driving a Predictive Text Entry System. *Frontiers in neuroscience*, 5:99, jan 2011. ISSN 1662-453X. doi: 10.3389/fnins.2011.00099. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3163907{&}tool=pmcentrez{&}rendertype=abstract. 93, 103, 104, 106

Johannes Höhne, Konrad Krenzlin, Sven Dähne, and Michael Tangermann. Natural stimuli improve auditory BCIs with respect to ergonomics and performance. *Journal of neural engineering*, 9(4), aug 2012. ISSN 1741-2552. doi: 10.1088/1741-2560/9/4/045003. URL http://iopscience.iop.org/article/10.1088/1741-2560/9/4/045003. 93, 103, 104, 106

Robert Horlings, Dragos Datcu, and L.J.M. Rothkrantz. Emotion recognition using brain activity. *Proceedings of the 9th International Conference on Computer Systems and Technologies and Workshop for PhD Students in Computing*, (March):6, 2008. doi: 10.1145/1500879.1500888. URL http://dl.acm.org/citation.cfm?id=1500888. 81

Anthony Hornof. The Prospects For Eye-Controlled Musical Performance. *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 461–466, 2014. URL http://www.nime.org/proceedings/2014/nime2014{_}562.pdf. 42

Anthony J. Hornof and Kyle E. V. Vessey. The Sound of One Eye Clapping: Tapping an Accurate Rhythm With Eye Movements. In *Proceedings of the Human Factors and Ergonomics Society*, volume 55, pages 1225–1229, 2011. doi: 10.1177/1071181311551255. 48, 51, 57

Anke Huckauf and Mario H. Urbina. Gazing with pEYEs. *Proceedings of the 2008 symposium on Eye tracking research & applications - ETRA '08*, pages 51–54, 2008. doi: 10.1145/1344471.1344483. URL http://dl.acm.org/citation.cfm?id=1344471.1344483$\delimiter"026E30F$nhttp://portal.acm.org/citation.cfm?doid=1344471.1344483. 42, 46

Andy Hunt, M.˜M. Marcelo M.˜M. Wanderley, and Matthew Paradis. The Importance of Parameter Mapping in Electronic Instrument Design. *Proceedings of the 2002 Conference on New Interfaces for Musical Expression*, 32(4):149–154, 2002. ISSN 0929-8215. doi: 10.1076/jnmr.32.4.429.18853. URL http://portal.acm.org/citation.cfm?id=1085207$\delimiter"026E30F$nhttp://www.nime.org/2004/NIME02/hunt.pdf. 58, 64, 68

K L Hyde, J Lerch, A Norton, M Forgeard, E Winner, A C Evans, and G Schlaug. Musical training shapes structural brain development. *Journal of Neuroscience*, 29(10):3019–3025, 2009. ISSN 0270-6474. doi: 10.1523/JNEUROSCI.5118-08.2009. URL http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve{&}db=PubMed{&}dopt=Citation{&}list{_}uids=19279238. 2

Robert J. K. Jacob. The use of eye movements in human-computer interaction techniques: what you look at is what you get. *ACM Transactions on Information Systems*, 9(3):152–169, 1991. ISSN 10468188. doi: 10.1145/123078.128728. 42

Yongwoong Jeon, Chang S. Nam, Young Joo Kim, and Min Cheol Whang.

Event-related (De)synchronization (ERD/ERS) during motor imagery tasks: Implications for brain-computer interfaces. *International Journal of Industrial Ergonomics*, 41(5):428–436, 2011. ISSN 01698141. doi: 10.1016/j.ergon.2011.03.005. URL http://dx.doi.org/10.1016/j.ergon.2011.03.005. 79

Sergi Jordà. The reactable. *Proceedings of the 28th of the international conference extended abstracts on Human factors in computing systems - CHI EA '10*, page 2989, 2010. doi: 10.1145/1753846.1753903. URL http://dl.acm.org/citation.cfm?id=1753846.1753903. 76

Patrik N Juslin. Communicating emotion in music performance: A review and a theoretical framework. In P N Juslin J A Sloboda, editor, *Music and emotion: Theory and research*, Series in affective science., pages 309–337. Oxford University Press, Juslin, Patrik N.: Uppsala U, Dept of Psychology, Box 1225, Uppsala, Sweden, SE - 751 42, 2001. ISBN 0-19-263189-6 (Hardcover); 0-19-263188-8 (Paperback). 81

Ivo Käthner, Carolin A Ruf, Emanuele Pasqualotto, Christoph Braun, Niels Birbaumer, and Sebastian Halder. A portable auditory P300 brain-computer interface with directional cues. *Clinical neurophysiology : official journal of the International Federation of Clinical Neurophysiology*, 124(2):327–38, feb 2013. ISSN 1872-8952. doi: 10.1016/j.clinph.2012.08.006. URL http://www.sciencedirect.com/science/article/pii/S1388245712005640. 93, 103, 117

Mandeep Kaur, P. Ahmed, and M. Qasim Rafiq. Technology development for unblessed people using bci: A survey. *International Journal of Computer Applications*, 40(1), February 2012. 71

Matthew C. Kiernan, Steve Vucic, Benjamin C. Cheah, Martin R. Turner, Andrew Eisen, Orla Hardiman, James R. Burrell, and Margaret C. Zoing. Amyotrophic lateral sclerosis. *The Lancet*, 377(9769):942–955, 2011. ISSN 01406736. doi: 10.1016/S0140-6736(10)61156-7. URL http://dx.doi.org/10.1016/S0140-6736(10)61156-7. 68

R Kirk, M Abbotson, R Abbotson, A Hunt, and A Cleaton. Computer music in the service of music therapy: the MIDIGRID and MIDICREATOR systems. *Medical engineering & physics*, 16(3):253–8, may 1994. ISSN 1350-4533. URL http://www.ncbi.nlm.nih.gov/pubmed/8061913. 9

D S Klobassa, T M Vaughan, P Brunner, N E Schwartz, J R Wolpaw, C Neuper, and E W Sellers. Toward a high-throughput auditory P300-based brain-computer interface. *Clinical neurophysiology : official journal of the International Federation of Clinical Neurophysiology*,

120(7):1252–61, jul 2009. ISSN 1872-8952. doi: 10.1016/j.clinph.2009. 04.019. URL http://www.sciencedirect.com/science/article/pii/ S1388245709003344. 92, 103, 106

Richard J Krauzlis. The Neuroscientist The Control of Voluntary Eye Movements :. *The Neuroscientist*, pages 124–137, 2005. ISSN 1073-8584. doi: 10.1177/1073858404271196. 41

Andrea Kübler, Nicola Neumann, Barbara Wilhelm, Thilo Hinterberger, and Niels Birbaumer. Predictability of Brain-Computer Communication. *Journal of Psychophysiology*, 18(2/3):121–129, jan 2004. ISSN 0269-8803. doi: 10.1027/0269-8803.18.23.121. URL http://econtent. hogrefe.com/doi/abs/10.1027/0269-8803.18.23.121. 100

Manu Kumar, Andreas Paepcke, and Terry Winograd. EyePoint. *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '07*, page 421, 2007. doi: 10.1145/1240624.1240692. URL http://portal.acm.org/citation.cfm?doid=1240624.1240692. 53

Bram Van De Laar, Ivo Brugman, Femke Nijboer, Mannes Poel, Anton Nijholt, and Bram Van De Laar. BrainBrush, a Multimodal Application for Creative Expressivity. *ACHI 2013 : The Sixth International Conference on Advances in Computer-Human Interactions*, (c):62–67, 2013. 117

Andrea Lamont, Roger Knox, Tom Chau, Yani Hamdani, Heidi Schwellnus, Cynthia Tam, and Patricia Johnson. Converting Movements to Music: New Musical Exploration Opportunities for Children in Rehabilitation. In *Proceedings of the 29th Annual Conference of the Canadian Association for Music Therapy*, number JANUARY, pages 26–31, 2000. 10

PJ Lang, MM Bradley, and BN Culthbert. International affective digitized sounds (iads): Stimuli, instruction manual and affective ratings (tech. rep. no. b-2). *The Center for Research in Psychophysiology, University of Florida, USA*, 1999. 83

Gustav Larsson. Evaluation Methodology of Eye Movement Classification Algorithms. pages 1–56, 2010. 53

TIFFANY LAWYER. AMYOTROPHIC LATERAL SCLEROSIS. *A.M.A. Archives of Neurology & Psychiatry*, 69(2):171, feb 1953. ISSN 0096-6886. doi: 10.1001/archneurpsyc.1953.02320260029002. URL http://archneurpsyc.jamanetwork.com.sare.upf.edu/article. aspx?articleid=651523. 68

Eui Chul Lee, Jin Cheol Woo, Jong Hwa Kim, Mincheol Whang, and Kang Ryoung Park. A brain-computer interface method combined with eye tracking for 3D interaction. *Journal of Neuroscience Methods*, 190

(2):289–298, 2010. ISSN 01650270. doi: 10.1016/j.jneumeth.2010.05.008. 73

Alexander Lenhardt, Matthias Kaper, and Helge J. Ritter. An adaptive P300-based online brain-computer interface. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 16(2):121–130, 2008. ISSN 15344320. doi: 10.1109/TNSRE.2007.912816. 91

Y. P. Lin, C. H. Wang, T. P. Jung, T. L. Wu, S. K. Jeng, J. R. Duann, and J. H. Chen. Eeg-based emotion recognition in music listening. *IEEE Transactions on Biomedical Engineering*, 57(7):1798–1806, July 2010. ISSN 0018-9294. doi: 10.1109/TBME.2010.2048568. 82, 88

Javier Lopez-Calderon and Steven J Luck. ERPLAB: an open-source toolbox for the analysis of event-related potentials. *Frontiers in human neuroscience*, 8(April):213, 2014. ISSN 1662-5161. doi: 10.3389/fnhum.2014.00213. URL http://journal.frontiersin.org/article/10.3389/fnhum.2014.00213/abstract. 111

Alvin Lucier. Statement on: music for solo performer. *Biofeedback and the Arts, Results of Early Experiments. Vancouver: Aesthetic Research Center of Canada Publications*, pages 60–61, 1976. 74

Päivi Majaranta and Andreas Bulling. Advances in Physiological Computing. *Advances in Physiological Computing*, pages 17–39, 2014. doi: 10.1007/978-1-4471-6392-3. URL http://link.springer.com/10.1007/978-1-4471-6392-3. 65

Pivi Majaranta and Kari-Jouko Rih. {CHAPTER} 9 - text entry by gaze: Utilizing eye tracking. In I. Scott MacKenzie and Kumiko Tanaka-Ishii, editors, *Text Entry Systems*, pages 175 – 187. Morgan Kaufmann, Burlington, 2007. ISBN 978-0-12-373591-1. doi: http://dx.doi.org/10.1016/B978-012373591-1/50009-7. URL http://www.sciencedirect.com/science/article/pii/B9780123735911500097. 65

Scott Makeig, Grace Leslie, Tim Mullen, Devpratim Sarma, Nima Bigdely-Shamlo, and Christian Kothe. First demonstration of a musical emotion bci. In Sidney DMello, Arthur Graesser, Bjrn Schuller, and Jean-Claude Martin, editors, *Affective Computing and Intelligent Interaction*, volume 6975 of *Lecture Notes in Computer Science*, pages 487–496. Springer Berlin / Heidelberg, 2011. ISBN 978-3-642-24570-1. URL http://dx.doi.org/10.1007/978-3-642-24571-8_61. 10.1007/978-3-642-24571-8_61. 77

Susana Martinez-Conde. Chapter 8 Fixational eye movements in normal and pathological vision. *Progress in Brain Research*, 154(SUPPL. A):

151–176, 2006. ISSN 00796123. doi: 10.1016/S0079-6123(06)54008-7. 41

Karl A. McCreadie, Damien H. Coyle, and Girijesh Prasad. Sensorimotor learning with stereo auditory feedback for a brain-computer interface. *Medical and Biological Engineering and Computing*, 51(3):285–293, 2013. ISSN 01400118. doi: 10.1007/s11517-012-0992-7. 78

Dennis J. McFarland and Jonathan R. Wolpaw. Brain-computer interfaces for communication and control. *Commun. ACM*, 54(5):60–66, May 2011. ISSN 0001-0782. doi: 10.1145/1941487.1941506. URL http://doi.acm.org/10.1145/1941487.1941506. 71

S Mealla, A Väljamäe, M Bosi, and S Jordà. Sonification of Brain and Body Signals in Collaborative Tasks Using a Tabletop Musical Interface. *Proceedings of 17th International Conference on Auditory Display (ICAD)*, pages 1–5, 2011. 75

Jrgen Mellinger, Gerwin Schalk, Christoph Braun, Hubert Preissl, Wolfgang Rosenstiel, Niels Birbaumer, and Andrea Kbler. An meg-based braincomputer interface (bci). *NeuroImage*, 36(3):581 – 593, 2007. ISSN 1053-8119. doi: 10.1016/j.neuroimage.2007.03.019. URL http://www.sciencedirect.com/science/article/pii/S1053811907002261. 71

Darius Miniotas. Application of Fitts' law to eye gaze interaction. *CHI '00 extended abstracts on Human factors in computer systems - CHI '00*, pages 339–340, 2000. doi: 10.1145/633482.633496. URL http://portal.acm.org/citation.cfm?doid=633292.633496. 47

Eduardo R. Miranda, Simon Durrant, and Torsten Anders. Towards braincomputer music interfaces: Progress and challenges. *2008 1st International Symposium on Applied Sciences in Biomedical and Communication Technologies, ISABEL 2008*, 44(0):1–5, 2008. doi: 10.1109/ISABEL.2008.4712626. 77

Eduardo Reck Miranda and Bram Boskamp. Steering Generative Rules with the EEG An Approach to Brain-Computer Music Interfacing. *Proceedings of Sound and Music Computing 05*, (Figure 1), 2005. 77

Eduardo Reck Miranda, Ken Sharman, Kerry Kilborn, and Alexander Duncan. On harnessing the electroencephalogram for the musical braincap. *Computer Music Journal*, 27(2), 2003. 76

Emily M. Mugler, Carolin A. Ruf, Sebastian Halder, Michael Bensch, and Andrea Kübler. Design and implementation of a P300-based braincomputer interface for controlling an internet browser. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 18(6):599–609, 2010. ISSN 15344320. doi: 10.1109/TNSRE.2010.2068059. 117

Toshimitsu Musha, Yuniko Terasaki, Hasnine A Haque, and George A Ivamitsky. Feature extraction from eegs associated with emotions. *Artificial Life and Robotics*, 1(1):15–19, 1997. 81

H Neville, Annika Andersson, Olivia Bagdade, and T. Effects of music training on brain and cognitive development in underpriviledged 3-to 5-year old children: Preliminary results. *Learning, Arts, and the Brain The Dana Consortium Report on Arts and Cognition*, pages 1–13, 2008. URL http://bdl.purpled.com/Publications/NevilleDanamonograph2008.pdf. 2

Ernst Niedermeyer and FH Lopes da Silva. *Electroencephalography: basic principles, clinical applications, and related fields*. Lippincott Williams & Wilkins, 2005. 71

Christopher P Niemic, Advisor Kirk, Warren Brown, and D Ph. Studies of Emotion: A Theoretical and Emperical Review of Psychophysiological Studies of Emotion. *Journal of Undergraduate Research*, pages 15–18, 2002. 83, 84

F. Nijboer, E. W. Sellers, J. Mellinger, M. A. Jordan, T. Matuz, A. Furdea, S. Halder, U. Mochty, D. J. Krusienski, T. M. Vaughan, J. R. Wolpaw, N. Birbaumer, and A. Kübler. A P300-based brain-computer interface for people with amyotrophic lateral sclerosis. *Clinical Neurophysiology*, 119 (8):1909–1916, 2008a. ISSN 13882457. doi: 10.1016/j.clinph.2008.03.034. 91

Femke Nijboer, Adrian Furdea, Ingo Gunst, Jrgen Mellinger, Dennis J. McFarland, Niels Birbaumer, and Andrea Kbler. An auditory braincomputer interface (bci). *Journal of Neuroscience Methods*, 167(1):43 – 50, 2008b. ISSN 0165-0270. doi: 10. 1016/j.jneumeth.2007.02.009. URL http://www.sciencedirect.com/science/article/pii/S0165027007000726. ¡ce:title¿Brain-Computer Interfaces (BCIs)¡/ce:title¿. 73, 75

Femke Nijboer, Bram van de Laar, Steven Gerritsen, Anton Nijholt, and Mannes Poel. Usability of Three Electroencephalogram Headsets for BrainComputer Interfaces: A Within Subject Comparison. *Interacting with Computers*, 27(5):500–511, sep 2015. ISSN 0953-5438. doi: 10.1093/iwc/iwv023. URL http://iwc.oxfordjournals.org/content/27/5/500.abstract. 94

Paul L Nunez and Ramesh Srinivasan. *Electric fields of the brain: the neurophysics of EEG*. Oxford University Press, USA, 2006. 72

T. Ohno. Features of eye gaze interface for selection tasks. In *Proceed-*

ings. *3rd Asia Pacific Computer Human Interaction (Cat. No.98EX110)*, pages 176–181. IEEE Comput. Soc, 1998. ISBN 0-8186-8347-3. doi: 10.1109/APCHI.1998.704190. URL http://ieeexplore.ieee.org/articleDetails.jsp?arnumber=704190. 42, 47

Pauline Oliveros, Leaf Miller, Jacklyn Heyen, Gillian Siddall, and Sergio Hazard. A musical improvisation interface for people with severe physical disabilities. *Music and Medicine*, 3(3):172–181, 2011. 10

Sile O'Modhrain. A Framework for the Evaluation of Digital Musical Instruments. *Computer Music Journal*, 35(1):28–42, 2011. ISSN 0148-9267. doi: 10.1162/COMJ{_}a{_}00038. 58

Giuseppe Pagnoni and Milos Cekic. Age effects on gray matter volume and attentional performance in Zen meditation. *Neurobiology of Aging*, 28(10):1623–1627, 2007. ISSN 01974580. doi: 10.1016/j.neurobiolaging.2007.06.008. 2

Robert Palisano, Peter Rosenbaum, Stephen Walter, Dianne Russell, Ellen Wood, and Barbara Galuppi. Development and reliability of a system to classify gross motor function in children with cerebral palsy. *Developmental Medicine and Child Neurology*, 39(2):214–223, 1997. ISSN 0012-1622. doi: 10.1111/j.1469-8749.1997.tb07414.x. 29

Alexandra Parbery-Clark, Dana L. Strait, Samira Anderson, Emily Hittner, and Nina Kraus. Musical experience and the aging auditory system: Implications for cognitive abilities and hearing speech in noise. *PLoS ONE*, 6(5), 2011. ISSN 19326203. doi: 10.1371/journal.pone.0018082. 2

G. Pfurtscheller, C. Brunner, A. Schl??gl, and F. H. Lopes da Silva. Mu rhythm (de)synchronization and EEG single-trial classification of different motor imagery tasks. *NeuroImage*, 31(1):153–159, 2006. ISSN 10538119. doi: 10.1016/j.neuroimage.2005.12.003. 79

F Piccione, F Giorgi, P Tonin, K Priftis, S Giove, S Silvoni, G Palmas, and F Beverina. P300-based brain computer interface: reliability and performance in healthy and paralysed participants. *Clinical neurophysiology : official journal of the International Federation of Clinical Neurophysiology*, 117(3):531–7, mar 2006. ISSN 1388-2457. doi: 10.1016/j.clinph.2005.07.024. URL http://www.ncbi.nlm.nih.gov/pubmed/16458069. 91

John R Pierce. *An introduction to information theory: symbols, signals and noise*. Courier Corporation, 2012. 95

John Polich. Updating P300: an integrative theory of P3a and P3b. *Clinical neurophysiology : official journal of the International Federation of Clinical Neurophysiology*, 118(10):2128–48, oct

2007. ISSN 1388-2457. doi: 10.1016/j.clinph.2007.04.019. URL http://www.ncbi.nlm.nih.gov/pubmed/17573239http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC2715154. 91

Rafael Ramirez and Zacharias Vamvakousis. Detecting emotion from EEG signals using the Emotive Epoc device. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 7670 LNAI:175–184, 2012. ISSN 03029743. doi: 10.1007/978-3-642-35139-6_17. 82

Brice Rebsamen, Etienne Burdet, Cuntai Guan, Haihong Zhang, Chee Leong Teo, Qiang Zeng, Christian Laugier, and Marcelo H. Ang Jr. Controlling a Wheelchair Indoors Using Thought. *IEEE Intelligent Systems*, 22(2):18–24, mar 2007. ISSN 1541-1672. doi: 10.1109/MIS.2007.26. URL http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4136854. 117

Stuart Reeves, Steve Benford, Claire O'Malley, and Mike Fraser. Designing the Spectator Experience. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2005:741–750, 2005. ISSN 1-58113-998-5. doi: 10.1145/1054972.1055074. URL http://eprints.nottingham.ac.uk/252/. 58

Yann Renard, Fabien Lotte, Guillaume Gibert, Marco Congedo, Emmanuel Maby, Vincent Delannoy, Olivier Bertrand, and Anatole Lécuyer. OpenViBE: An Open-Source Software Platform to Design, Test, and Use BrainComputer Interfaces in Real and Virtual Environments. *Presence: Teleoperators and Virtual Environments*, 19(1):35–53, feb 2010. ISSN 1054-7460. doi: 10.1162/pres.19.1.35. URL http://www.mitpressjournals.org/doi/abs/10.1162/pres.19.1.35{#}.VkNH9rcvdpg. 84, 86, 96, 108

Bruno H Repp. Sensorimotor synchronization: a review of the tapping literature. *Psychonomic bulletin & review*, 12(6):969–92, dec 2005. ISSN 1069-9384. URL http://www.ncbi.nlm.nih.gov/pubmed/16615317. 48

Bertrand Rivet, Antoine Souloumiac, Virginie Attina, and Guillaume Gibert. xDAWN algorithm to enhance evoked potentials: application to brain-computer interface. *IEEE transactions on bio-medical engineering*, 56(8):2035–43, aug 2009. ISSN 1558-2531. doi: 10.1109/TBME.2009.2012869. URL http://www.ncbi.nlm.nih.gov/pubmed/19174332. 98, 120

Peter Rosenbaum, Nigel Paneth, Alan Leviton, Murray Goldstein, Martin Bax, Diane Damiano, Bernard Dan, and Bo Jacobsson. A re-

port: The definition and classification of cerebral palsy April 2006. *Developmental Medicine and Child Neurology*, 49(SUPPL.109):8–14, feb 2007. ISSN 00121622. doi: 10.1111/j.1469-8749.2007.tb12610.x. URL http://www.ncbi.nlm.nih.gov/pubmed/17370477. 29

D. D. Salvucci and J. H. Goldberg. Identifying Fixations and Saccades in Eye-Tracking Protocols. *Proceedings of the Eye Tracking Research and Applications Symposium*, pages 71–78, 2000. ISSN 10960384. doi: 10.1145/355017.355028. 53

J San Agustin. Off-the-Shelf Gaze Interaction. *Communication*, (September):179, 2010. URL http://scholar.google.com/scholar?hl=en{&}btnG=Search{&}q=intitle:Off-the-Shelf+Gaze+Interaction{#}3. 53

Joan Sandoval Codina. *Tocando la guitarra con potenciómetros rotatorios: un nuevo controlador de guitarra diseñado para una persona con capacidades sensomotoras reducidas.* PhD thesis, Universitat Pompeu Fabra, dec 2015. URL http://repositori.upf.edu/handle/10230/25430. 21

Gottfried Schlaug, Lutz Jäncke, Yanxiong Huang, Jochen F. Staiger, and Helmuth Steinmetz. Increased corpus callosum size in musicians. *Neuropsychologia*, 33(8):1047–1055, 1995. ISSN 00283932. doi: 10.1016/0028-3932(95)00045-5. 2

W Andrew Schloss. Using Contemporary Technology in Live Performance : The Dilemma of the Performer *. *Journal of New Music Research*, 31 (February 2015):293–242, 2002. ISSN 0929-8215. doi: 10.1076/jnmr.32.3.239.16866. 58

Marcel Schmidt. *A Guitar Controller for People with Cerebral Palsy.* Master thesis, Universitat Pompeu Fabra, 2014. 29, 32

Bernhard Scholkopft and Klaus-Robert Mullert. Fisher discriminant analysis with kernels. *Neural networks for signal processing IX*, 1(1):1, 1999. 86

Martijn Schreuder, Benjamin Blankertz, and Michael Tangermann. A new auditory multi-class brain-computer interface paradigm: spatial hearing as an informative cue. *PloS one*, 5(4):e9813, jan 2010. ISSN 1932-6203. doi: 10.1371/journal.pone.0009813. URL http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0009813. 92, 93, 102, 103, 104, 106, 115

Martijn Schreuder, Thomas Rost, and Michael Tangermann. Listen, You are Writing! Speeding up Online Spelling with a Dynamic Auditory BCI. *Frontiers in neuroscience*, 5:112, jan

2011. ISSN 1662-453X. doi: 10.3389/fnins.2011.00112. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid= 3192990{&}tool=pmcentrez{&}rendertype=abstract. 93, 103, 104, 117

Eric W. Sellers and Emanuel Donchin. A P300-based brain-computer interface: Initial tests by ALS patients. *Clinical Neurophysiology*, 117(3): 538–548, 2006. ISSN 13882457. doi: 10.1016/j.clinph.2005.06.027. 91

Frederick Shic, Brian Scassellati, and K. Chawarska. The incomplete fixation measure. *Proceedings of the 2008 symposium on Eye tracking research & applications*, (Figure 1):111–114, 2008. doi: 10.1145/1344471.1344500. URL http://dl.acm. org/citation.cfm?id=1344500$\delimiter"026E30F$npapers3: //publication/uuid/D1676427-B5B8-4821-A73E-3934C37B642A$\ delimiter"026E30F$nhttp://www.cs.yale.edu/homes/scaz/papers/ Shic-ETRA-08.pdf. 53

Nadine Simon, Ivo Käthner, Carolin A Ruf, Emanuele Pasqualotto, Andrea Kübler, and Sebastian Halder. An auditory multiclass brain-computer interface with natural stimuli: Usability evaluation with healthy participants and a motor impaired end user. *Frontiers in human neuroscience*, 8:1039, jan 2014. ISSN 1662-5161. doi: 10.3389/fnhum.2014.01039. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid= 4288388{&}tool=pmcentrez{&}rendertype=abstract. 94, 103

Ranganatha Sitaram, Andrea Caria, Ralf Veit, Tilman Gaber, Giuseppina Rota, Andrea Kuebler, and Niels Birbaumer. fmri brain-computer interface: a tool for neuroscientific research and treatment. *Intell. Neuroscience*, 2007:1:1–1:10, April 2007. ISSN 1687-5265. doi: 10.1155/2007/ 25487. URL http://dx.doi.org/10.1155/2007/25487. 71

Vanessa Sluming, Thomas Barrick, Matthew Howard, Enis Cezayirli, Andrew Mayes, and Neil Roberts. Voxel-Based Morphometry Reveals Increased Gray Matter Density in Broca's Area in Male Symphony Orchestra Musicians. *NeuroImage*, 17(3):1613–1622, 2002. ISSN 10538119. doi: 10.1006/nimg.2002.1288. URL http://www.sciencedirect.com/ science/article/pii/S1053811902912887. 2

Dana L. Strait, Nina Kraus, Alexandra Parbery-Clark, and Richard Ashley. Musical experience shapes top-down auditory mechanisms: Evidence from masking and auditory attention performance. *Hearing Research*, 261(1-2):22–29, 2010. ISSN 03785955. doi: 10.1016/j.heares.2009.12.021. URL http://dx.doi.org/10.1016/j.heares.2009.12.021. 2

Ben Swift, James Sheridan, Yang Zhen, and Henry J. Gardner. Mind-modulated music in the mind attention interface. In *Proceedings of the 19th Australasian conference on Computer-Human Interaction: Entertaining User Interfaces*, OZCHI '07, pages 83–86, New York, NY, USA, 2007. ACM. ISBN 978-1-59593-872-5. doi: 10.1145/1324892.1324907. URL http://doi.acm.org/10.1145/1324892.1324907. 76

Tim Swingler. *" That Was Me!": Applications of the Soundbeam MIDI Controller as a Key to Creative Communication, Learning, Independence and Joy.* ERIC, 1998. 10

K. Takahashi. Remarks on emotion recognition from multi-modal bio-potential signals. In *Industrial Technology, 2004. IEEE ICIT '04. 2004 IEEE International Conference on*, volume 3, pages 1138–1143 Vol. 3, Dec 2004. doi: 10.1109/ICIT.2004.1490720. 82

Cynthia Tam, Heidi Schwellnus, Ceilidh Eaton, Yani Hamdani, Andrea Lamont, and Tom Chau. Movement-to-music computer technology: a developmental play experience for children with severe physical disabilities. *Occupational Therapy International*, 14(2):99–112, 2007. 10

Wolfgang A. Teder-Sälejärvi and Steven A. Hillyard. The gradient of spatial auditory attention in free field: An event-related potential study. *Perception & Psychophysics*, 60(7):1228–1242, jan 1998. ISSN 0031-5117. doi: 10.3758/BF03206172. URL http://www.springerlink.com/index/10.3758/BF03206172. 92

Richard Teitelbaum. In tune: Some early experiments in biofeedback music (1966-1974). *Biofeedback and the Arts, Results of Early Experiments. Vancouver: Aesthetic Research Center of Canada Publications*, 1976. 74

A. Väljamäe, S. Mealla, Tony Steffert, S. Holland, X. Marimon, R. Benitez, A. Oliveira, and S. Jordà. A review of real-time EEG sonification research. *International Conference on Auditory Display*, (July):85–93, 2013. 74

Z Vamvakousis and R Ramirez. Towards a Low Cost Mu-Rhythm Based BCI. In *Proceedings of the Fifth International Brain-Computer Interface Meeting 2013*, 2013. ISBN 9783851252606. doi: 10.3217/978-3-85125-260-6-13. 79

Zacharias Vamvakousis and Rafael Ramirez. *The eyeharp: A gaze-controlled musical instrument.* Master thesis, Universitat Pompeu Fabra, 2011. URL http://mtg.upf.edu/system/files/publications/Vamvakousis-Zacharias-Master-thesis-2011.pdf. 44, 90

Zacharias Vamvakousis and Rafael Ramirez. Temporal control in the Eye-Harp gaze-controlled musical interface. In Georg Essl, Brent Gillespie,

Michael Gurevich, and Sile O'Modhrain, editors, *Proceedings of the 12th International Conference on NIME*, pages 11–16. Division of Computer Science & Engineering,, 2012. 49

Zacharias Vamvakousis and Rafael Ramirez. P300 harmonies: A Brain-Computer Musical Interface. *Proceedings - 40th International Computer Music Conference, ICMC 2014 and 11th Sound and Music Computing Conference, SMC 2014 - Music Technology Meets Philosophy: From Digital Echos to Virtual Ethos*, pages 725–729, 2014. URL http://www.scopus.com/inward/record.url?eid=2-s2.0-84908883844{&}partnerID=tZOtx3y1. 118

Zacharias Vamvakousis and Rafael Ramirez. Is an auditory P300-based Brain-Computer Musical Interface feasible ? In *1st International BCMI Workshop*, pages 1–8, 2015. 121

Zacharias Vamvakousis and Rafael Ramirez. The eyeharp: A gaze-controlled digital musical instrument. *Frontiers in Psychology*, 7 (906), 2016. ISSN 1664-1078. doi: 10.3389/fpsyg.2016.00906. URL http://www.frontiersin.org/performance_science/10.3389/fpsyg.2016.00906/abstract. 58

M. D. van der Zwaag, J. H. D. M. Westerink, and E. L. van den Broek. Emotional and psychophysiological responses to tempo, mode, and percussiveness. *Musicae Scientiae*, 15(2):250–269, 2011. ISSN 1029-8649. doi: 10.1177/1029864911403364. 90

M Vourkas, S Micheloyannis, and G Papadourakis. Use of ann and hjorth parameters in mental-task discrimination. In *Advances in medical signal and information processing, 2000. First international conference on (IEE conf. publ. no. 476)*, pages 327–332. IET, 2000. 81

Catherine Y Wan and Gottfried Schlaug. Music making as a tool for promoting brain plasticity across the life span. *The Neuroscientist : a review journal bringing neurobiology, neurology and psychiatry*, 16(5): 566–77, 2010. ISSN 1089-4098. doi: 10.1177/1073858410377805. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2996135{&}tool=pmcentrez{&}rendertype=abstract. 2

Dongcui Wang, Zhencheng Chen, Chao Yang, Jun Liu, Fongming Mo, and Yangde Zhang. Validation of the Mobile Emotiv Device Using a Neuroscan Event-Related Potential System. *Journal of Medical Imaging and Health Informatics*, 5(7):1553–1557, dec 2015. ISSN 21567018. doi: 10.1166/jmihi.2015.1563. URL http://www.ingentaconnect.com/content/asp/jmihi/2015/00000005/00000007/art00031. 94, 107

Yijun Wang, Xiaorong Gao, Bo Hong, and Shangkai Gao. Practical Designs of BrainComputer Interfaces Based on the Modulation of EEG Rhythms. pages 137–154. Springer Berlin Heidelberg, 2009. doi: 10.1007/978-3-642-02091-9_8. URL http://link.springer.com/10.1007/978-3-642-02091-9{_}8. 79

Jonathan R. Wolpaw, Dennis J. McFarland, Gregory W. Neat, and Catherine A. Forneris. An EEG-based brain-computer interface for cursor control. *Electroencephalography and Clinical Neurophysiology*, 78(3):252–259, 1991. ISSN 00134694. doi: 10.1016/0013-4694(91)90040-B. 78

Jonathan R Wolpaw, Niels Birbaumer, Dennis J McFarland, Gert Pfurtscheller, and Theresa M Vaughan. Braincomputer interfaces for communication and control. *Clinical Neurophysiology*, 113(6): 767 – 791, 2002. ISSN 1388-2457. doi: 10.1016/S1388-2457(02) 00057-3. URL http://www.sciencedirect.com/science/article/pii/S1388245702000573. 71

Dan Wu, Yang Liu, and De-zhong Yao. Listen to the Brain in Real Time The Chengdu Brainwave Music. *Science And Technology*, 7(1):34–39, 2009. 75

Alfred L. Yarbus. Eye movements and vision. *Neuropsychologia*, 6(4):222, 1967. ISSN 00283932. doi: 10.1016/0028-3932(68)90012-2. 41

Thorsten O. Zander, Matti Gaertner, Christian Kothe, and Roman Vilimek. Combining eye gaze input with a braincomputer interface for touchless humancomputer interaction. *International Journal of Human-Computer Interaction*, 27(1):38–51, 2010. doi: 10.1080/10447318.2011.535752. URL http://www.tandfonline.com/doi/abs/10.1080/10447318.2011.535752. 72