TUFTS UNIVERSITY

An Adaptive fNIRS-based BCI for Learning Music on the Piano

by

Kurt Oleson

Senior Honors Thesis

in the
Department of Human Computer Interaction
Committee: Robert Jacob, Aniruddh Patel, Paul Lehrman

August 2015

"The function of music is to release us from the tyranny of conscious thought."

Sir Thomas Beecham (1879-1961)

TUFTS UNIVERSITY

Abstract

Department of Human Computer Interaction

Senior Honors Thesis

by Kurt Oleson

The brain is constantly working and adapting to the environments that we find ourselves in, why shouldnt the tasks that we perform adapt as well? Brain computer interfaces have been used to assist people with disabilities as well as to provide passive information in order to simplify tasks for a user. This experiment sought to utilize an fNIRS-based BCI in a learning environment. The interface tested was aimed to improve learning speed and accuracy. Specifically, this study focused on learning a piece of music on the piano. The BCI used cognitive load measurements in order to adjust the difficulty of music presented to participants in real time. Our findings showed that participants improved in objective performance measures and reported a better learning and understanding of the music when learning with the BCI as opposed to the control condition. These findings suggest that a passive BCI measuring cognitive load can be effectively used to improve learning for beginner pianists and that managing cognitive load in a learning environment can have performance benefits.

Acknowledgements

Thank you to everyone who has helped me complete this project. Thank you to my friends who have seen me put countless hours into this project and have supported me through some difficult times; special thanks go to Joey, Grace, Joel, Cami, Chris, Kristen and Abby. Thanks to my family for constant support and encouragement, I wouldn't have gotten this far without Mom and Dad, they really don't get enough credit. Thanks to the other members of the HCI lab, Dan and Remco, who were instrumental in guiding our work to it's final iteration; special thanks to Lane for working through the statistics with us. Thank you to my advisers Ani Patel and Paul Lehrman for their feedback, advice, and confidence in my abilities. Thank you to my thesis adviser Rob Jacob who has been a proponent of my involvement in the lab since day one and has provided invaluable advice in the design and direction of the project. Finally, I would like to thank Beste Yuksel who is not only the lead on this project but has been a great friend and mentor. She challenges my ideas, forces me to think critically about each aspect of the study, and is the backbone of this project. If it weren't for her I wouldn't be writing this thesis. Collaborating with Beste has greatly impacted my future and I am very grateful to have worked with her.

Contents

List of Figures

List of Tables

Abbreviations

BCI	**B**rain **C**omputer **I**nterface
fNIRS	**f**unctional **N**ear-**I**nfrared **S**pectroscopy
MIDI	**M**usical **I**nstrument **D**igital **I**nterface
HCI	**H**uman **C**omputer **I**nteraction
DAW	**D**igital **A**udio **W**orkstation
BPM	**B**eats **P**er **M**inute

Chapter 1

Introduction

The power of the human brain is greatly underestimated and for the large part still a mystery. We are constantly processing information and are often unaware of it. In the emerging field of brain computer interfaces, we are getting closer to harnessing the information from these unconscious processes and applying it to tasks in our everyday lives. If we can measure someones brain activity then why not apply that technology to the task they are focusing on? With adaptive brain computer interfaces there are possibilities to increase performance, multitasking, and efficiency.

Every brain is unique and the tools we have to measure brain activity are still limited in effectiveness and accessibility. However, recent developments have improved our abilities to monitor and assess how hard the brain is working. With this in mind, we should consider how this activity can be used and balanced by different interfaces that take advantage of data produced by the brain.

For those of us who are musicians, we know that learning an instrument takes many years of practice. Furthermore, we know that even after learning to play an instrument reading music can be another source of difficulty. Even for those us who dont play an instrument, it is hard to process how written music is transformed in to the songs and sounds that we hear. If only there was some way to assist in learning to read a piece of music that didn't make the musician work any harder and that could even facilitate the process. What if we could use our brains to somehow augment learning music? In this paper I will propose a system for accomplishing such a goal as well as outline the design and results of an experiment that was run to test the effectiveness of such an approach.

As a musician and scientist, the intersection between the two has always been of great interest to me. For this project, I spearheaded all of the musical data acquisition and analysis using a combination of manual and automatic assessment that tested my abilities as an engineer, researcher, and musician. Chapter 4 is where I outline my largest contributions to this project.

The motivation for this project was largely driven by Yuksel et al.'s previous work [1] measuring cognitive workload in pianists during musical improvisation. We wanted to adapt this project to apply to an objective measure of cognitive workload to learning. More generally, we sought to use cognitive workload as a variable in a learning domain as there are very few examples of this in existing literature. Ultimately our goal was to use a passive measurement of cognitive load to improve learning in a musical context.

Chapter 2

Related Work

2.1 BCI Background

In a technologically diverse and ever expanding society, there is increasing opportunity to use more advanced ways to interact with the world around us. One emerging available method is the use of Brain Computer Interfaces (BCI). These interfaces have been used in many ways [2] and the passive method of BCI implementation has become more popular in HCI research in recent years.

Prior to the idea of passive input, BCIs were used for explicit controlling systems for individuals who may have been paralyzed or otherwise impaired and had to use brain activity to control their surroundings. For healthy users this is usually not a very useful method of interaction as there is usually a serious training period and it is often slower than normal methods of input or interaction.

The goal of a passive BCI is to provide some sort of feedback or control to a program or system where the user does not actively try to alter their brain activity in order to receive some response. As opposed to measuring brain activity that a user is attempting to control, the BCI gathers data about the user's overall cognitive state. From this data, some parameter of a system can be altered based on certain "mental states" determined by the researcher.

There are many different methods of measuring brain activity for BCIs. The most commonly used methods are Electroencephalography (EEG) which measures electrical activity produced by the brain and functional Near Infrared Spectroscopy (fNIRS) which measures changes of oxygenated hemoglobin in blood due to brain activity. These are most commonly used because they have fairly good temporal resolution which is important when you are trying to control something in real time simply by thinking.

Most real-time BCIs can measure brain activity fairly accurately in both the temporal and spatial dimension. However, what does it mean when we see higher activation in the frontal lobe or an increased response in the P300 wave? Do these things show that we are cognitively overloaded/overwhelmed or that we are 'in the zone' and focusing well and efficiently? This is one of the biggest issues the BCI field is facing; what exactly are we seeing in the brain and how can we use it? This problem is pervasive and has not been solved, in this study we are focused on the measurement of a user's cognitive load and the prefrontal cortex has been shown to be linked to complex problem solving and multitasking. Since the fNIRS method reliably measures activation in this cortex we can posit that higher activation in this level means a higher load in problem solving and generally higher level cognitive processes and use this in our BCI.

2.2 Cognitive Load Theory

In this study, our BCI is designed to measure a user's brain activity and adjust parameters based on this data. In designing a BCI, it is important to identify not only what activity is being measured in the brain but how it relates to the task at hand. Those in the HCI field often refer to a "cognitive load" of participants. It is important to understand cognitive load with respect to this paper.

Cognitive load can mean many things but for the purpose of this paper will use the term as defined in what is known as Cognitive Load Theory (CLT) [3]. The theory consists of developing efficient uses of people's limited ability to process information in order to improve their learning. The key processes that make up CLT are working memory, acquired knowledge in long term memory, and learning [4]. When asked to learn new information, there are a number of factors that affect the efficiency and success of learning. First is working memory; there is only so much information that we can store. Second is previous knowledge of the information stored in long term memory. According to CLT, with enough prior knowledge of the presented information, one can better organize information in working memory. This, in theory, means the individual has a lower cognitive load, easier learning, and improved performance.

Since this study focuses on piano, let's use that as an example. For a skilled musician and pianist, seeing a page of sheet music for the first time will require some work and concentration to learn and play. But since they have prior knowledge of how to read music, they can more easily group the notes in their learning process. Compare this to a novice pianist with very little previous knowledge about reading music; they would likely have a harder time learning the piece and need to pay more attention to each individual piece of information. CLT tells us that the skilled pianist actually groups some of this information in a way so that the working memory is only doing the bare minimum whereas the novice does most of the processing in working memory [5]. Presumably, the novice will have a higher cognitive load while learning the piece than the skilled pianist.

It is clear to see that CLT is relevant in many contexts that have to do with learning or performance. Other work has been done with multitasking [6] and preference [7] and has shown fNIRS to be a reliable and effective measurement of cognitive load [8].

2.3 Applications to Learning

Most research conducted on cognitive load is in easily quantifiable, discrete contexts. Music is a very creative medium with little standardization. Creating an effective BCI for musicians required certain considerations based on past work.

The central assumption of theories such as CLT [3] and Baddeley's working memory model [9] is that the human system for processing higher-level information is finite. This limited cognitive capacity affects the learning of new tasks, where a learner can experience cognitive overload, in which the required cognitive processes exceeds the learner's available, finite cognitive capacity [10]. Cognitive overload is a central challenge when designing systems for learners because any kind of meaningful learning requires cognitive processing beyond learners' cognitive capacity [10].

During training tasks for the learner, part-task selection approaches have been used as segments or simplifications of the whole-task to aid users [11]. Dynamic task selection has been found to be superior to static task selection [12] as the training task can be adjusted to the cognitive state of the learner. However, most intelligent tutoring systems (ITS) do not take learner cognitive workload into account when adapting the selection of tasks, focusing instead on performance measures. Salden et al. [12] used cognitive workload as reported subjectively by learners in addition to performance measures to dynamically adapt learner training tasks but concluded that further work was needed in this area. Paas et al. [13] has also brought up the need to use cognitive workload as a measurement for learner training tasks and suggested brain sensing as possible alternative to subjective ratings by the learner.

In this work, we present an objective way of measuring cognitive workload using brain sensing, in a dynamic, part-task approach, and compare it with a control condition.

2.4 Music Applications of BCIs

While other BCIs aim to assist a user in some way or to make a task easier, musical BCIs have, for the most part, been aimed towards composition and performance; geared towards creative applications as opposed to practical ones [14–16]). What makes music such an interesting media to use with a BCI is its expressiveness and possibility to change at any time. Other performance tasks traditionally used with BCIs tend to be more discrete and do not allow for as much variation as playing an instrument does. For this reason, the musical BCI field is mostly focused on composition and manipulation of music in real time with brain data.

One of the few applications that has a more practical impact is one done by Sourina, Liu, and Nguyen in 2011 focusing on a real time music therapy BCI. By monitoring the individuals' brain signals, they assessed their emotional state and altered the music they heard accordingly; not for creative reasons but rather to improve the therapy that was being provided to the individual [17]. This study is a great example of how brain data can be used to passively change some musical parameters to benefit the user of the BCI. However, the benefit is not affecting the user's musical performance which is where research in this area is very seriously lacking.

The first half of the experiment discussed in this paper, was carried out and verified by Yuksel et al. [1] in the same lab with the same equipment and materials. The aim of this previous study, like others in its field, was to create a BCI that augmented a user's piano performance by adding different musical elements to an improvisation; another expressive interface named BRAAHMS. What was so novel about this experiment is that the BCI utilized fNIRS as a measure of cognitive load and altered the type and amount of musical additions based on how high and low the perceived cognitive load was. With this system, pianists were improvising while musical harmonies determined by pilot studies were added and removed based on user cognitive workload. Findings from this study showed that users preferred the BCI system while improvising in comparison to two control conditions because they felt more creative and that the system was responding to them [1]. Following this study, we came up for the design of the study outlined in this paper.

In the field of musical BCIs research has shown that EEG and fNIRS can both be used for musical composition and altering music in real time. These applications, however, do not give any direct benefit to the user other than an additional creative outlet. Following the Yuksel study, we decided to apply a BCI to a learning context within music. Instead of extending the expressive nature of music, we focused on a problem (learning written music) and set to solve in using passive brain input. Instead of focusing on using passive brain input to change or create the music the user is playing, our approach uses the brain measurement to alter the music notation being displayed to the user. One of the largest contributions from the BRAAHMS experiment was that it showed a clear distinction in brain activity when users were playing something difficult versus something easy. Based on this finding, we knew that it would be appropriate to use cognitive load measurements as a measure of difficulty.

As opposed to simply monitoring a user's cognitive load, our system, in theory, should help manage the cognitive load and prevent it from being too low or too high for extended periods of time. By adjusting the music notation in real time, the difficulty of the piece changes, thus affecting the user's cognitive load. While monitoring cognitive load, we can make predictions as to whether a user is ready for increased difficulty in music. Based on the study by Yuksel et al. [1] and pilot experiments we targeted beginner pianists in the hopes that this novel, adaptive BCI would improve learning accuracy, speed, and level of enjoyment. With this assumption in mind we thought it important to consider the applications of this study in the context of music education.

2.5 Music Education

There is seemingly endless research done on music education, especially with new technologies and ways for students to interact with instruments and teachers to interact with students. Much of the research in this field focuses on evaluating certain methods of music education such as multimedia [18], augmented reality [19], or sonification feedback [20]. In addition to these evaluation studies, work has been done to show long term effects of music education and instrument proficiency on the brain [21, 22]. However, there is little to no work done on how learning music is effecting the brain in real time or how brain activity effects the way that music is learned.

Learning music, especially as a beginner, can be a daunting task that often requires a great deal of concentration and mental effort. It is reasonable to assume that the brain of someone learning a piece of music for the first time is working very hard. Of course, this is something that music educators have always known, hence the plethora of research in different methods of music education. While they may not admit to it explicitly, teachers that come up with new methods of learning music are doing so in order to make things easier; to reduce the cognitive load of their students.

One study of interest, although a bit outdated, is one run by R. Brown that compared the hands apart and hands together methods, two methods of learning music that are still in common practice today. Her findings showed that practicing and playing with both hands together, as opposed to learning each part separately, was a more efficient method of learning music [23]. However, teachers still implement this method today and even students who aren't instructed to do so will follow this practice on their own will. Why do we do this? Well perhaps it has to do with our brain. Performing a task with one hand certainly seems like it would be less cognitively demanding than using both hands. In the context of piano and music education, there hasn't been any work to see what the brain is doing while it learns which may give some explanations for the effectiveness of certain practices and methods.

With our design we hope to fill some of the gaps in the music education literature. Particularly with respect to how brain activity changes and effects the speed and accuracy of someone learning a piece of piano. We hope to show not only beneficial effects from utilizing cognitive load measurements but also that there are noticeable changes in activity which we will do by constantly measuring the brain through each task of our experiment.

Chapter 3

Experimental Design

3.1 Research Goals

The primary goal of this experiment was to create and assess the effectiveness of an fNIRS-based BCI to assist beginner pianists in learning to play pieces of music. We theorized that altering the difficulty of the music based on measured cognitive load would improve the speed and accuracy of participants performance after learning a piece. Thus, our hypothesis was as follows:

- Hypothesis: We will observe better piano performances of pieces of music that were learned by participants using our BCI, indicating that passive brain input can be successfully used to improve learning music.

In the sections below I outline the materials and methods of the experiment, explain the technology used, outline the considerations of BCI design and discuss the iterations of the experiment.

3.2 Materials

3.2.1 Music for Training Task

For the training task music was chosen with varied difficulties. Work was done with a music faculty professor to choose pieces that would be considered difficult and create pieces that were easy to play. We chose 15 easy pieces and 15 hard pieces for participants to play on the piano for 30 seconds at a time. Figure 3.1 below shows some brief examples.

FIGURE 3.1: Two measures from a hard piece (left) and an easy piece (right) used in the training task)

Criteria for the "easy" pieces:

- All notes were in C major (i.e. no sharps (♯) or flats (♭) in the key signature)
- Only whole notes were used (𝅝) (slow, long notes)
- There were no accidentals - no additional sharps, flats, or naturals (♮) that were not part of the key signature
- All notes were within a five note range (C to G) so that hand movement was minimal
- There were no dynamic changes (i.e. changes in volume of a note or stylistic execution)

Criteria for the "difficult" pieces:

- All pieces were in a more difficult key signature (most pieces had a key signature of at least 3 sharps or flats)
- Pieces contained accidentals
- Pieces contained mostly eighth (♪) and sixteenth notes (𝅘𝅥𝅯) (short, fast notes)
- Music required some moving of the hands but not too excessively
- Music included dynamic changes

Figures B.2 and B.1 in section B.1 show full examples of an easy and a hard piece.

3.2.2 Music for Learning Task

For the learning task, two Bach chorales were chosen. These chorales were chosen because of their similarity in style and difficulty. It can often times be hard to find two different pieces of music that can objectively said to be similar in level of difficulty. These chorales, however, were composed using the principles of musical counterpoint. Counterpoint is a set of rules and guidelines that dictate how music can be composed. It is a relationship between the voices of a piece that is interdependent harmonically but independent in contour and rhythm. In other words, the two chorales that were chosen are standardized based on this principle. So while they are different pieces of music, they are similar enough so that the difficulty level is not noticeably different. Both pieces were transposed to the same key and are the same length. Each piece also has 4 voices bass, tenor, alto, and soprano. This allowed for an easy way to segment the music in the BCI condition in order to progressively increase the difficulty. Due to experimental constraints and skill level of participants, the pieces were slightly altered. They were transposed into the same key (G Major) and the eighth notes were removed so that it is rhythmically consistent. It is important to note that by removing some of the notes, some minor counterpoint rules are no longer met. However, the underlying structure and form of both pieces is still in tact. The pieces used in this study can be found in section B.2.

3.3 Participants

For this experiment we recruited 18-25 year-olds that considered themselves to be beginner piano players. We found participants through general advertisements for the study as well as through the music department at Tufts University. We had 14 participants ranging from age 19 to 21. There were 7 males and 7 females. All participants were paid $20 for the first visit and an additional $20 for a next day follow-up.

3.4 Methods

The first task that participants had to complete was a training task for our machine learning model. This task consisted of 30 short sections of sheet music of varying difficulties. The participants began with a 60 second baseline task during which they were resting while a cross was on a screen. This allowed the system to learn what their baseline brain activity was like. Following this, the pieces of music were randomly presented to the participant with 30 second rest periods in between. A cross was again presented on the screen during these rests. By the end of this task, the system constructed a model of what a participant's workload looks like while playing something difficult and while playing something easy. This model is what was used for the next part of the experiment. This task has been used in previous experiments.

Next was the learning task. In this task, the participants were given a piece of music to learn as best they could within a given time frame of 15 minutes. They then repeated this task with a different piece of music. In one condition, the music was presented as it normally would be as if reading it off the page . The other condition utilized our BCI to alter the music that they saw. The music (described in section 3.2.2) was presented one voice at a time. The participants first saw a single line of music and more lines were progressively added. Based on the model created in the first part of the experiment, the system constantly makes predictions of whether the participants cognitive load is low or high. If their cognitive load dropped below a certain threshold for a long enough period of time then the music was made more difficult with the addition of another voice/line of notes to the score that they are learning. At the end of each 15 minute learning period participants were given a short minute-long rest and then asked to play once through the piece they had just learned as best they could. Their performance was measured using software outlined in chapter 4.

After completing the learning task, participants filled out questionnaires pertaining to the pieces of music they had just learned. These questionnaires (seen in section C) were very short and consisted of a self evaluation of different aspects of the learning task and performance.

Participants were given the option to come in for a follow up study within the following 2 days of the first experiment. For the follow up participants are asked to play once through each piece in the same order that they learned them in the previous study. This was done in order to measure any retention effects of the BCI system. In this study there was no fNIRS device used and no training or learning phase, just a performance task.

This was a balanced, within subjects design. All participants completed the training task followed by both the normal and BCI condition of the learning task. The order of the learning task alternated for each participant to account for the possibility of a learning effect. Following the piano tasks, participants were asked to fill out questionnaires about their preferences and performance and were interviewed about their experience. The experimental set up can be seen in Figure 3.4.

3.5 Technology Used

3.5.1 fNIRS

We used a multichannel frequency domain Imagent fNIRS device from ISS Inc. (Champaign, IL) for our data acquisition. Two probes were used to measure activity from both hemispheres of the prefrontal cortex of participants. The probes were placed next to each other on the participant's forehead. Each probe contains a detector and four light sources with each one emitting near-infrared light at two wavelengths (690 and 830 nm). With this set up we had sixteen data channels (2 probes x 4 source-detector pairs x 2 wavelengths) (Figure 3.2). The source-detector distances ranged from 1.5 and 3.5 cm, and the sampling rate was 11.79 Hz. The signals were filtered for heart rate, respiration, and movement artifacts using a third-degree polynomial filter and low-pass elliptical filter. The fNIRS data was sent to another computer for processing by a custom system created by the HCI lab for past experiments.

FIGURE 3.2: fNIRS sensor with 5 light sources (only 4 sources were used for this study)

3.5.2 MATLAB and Imagent Computers

Training Task Modeling

For each user, we created a unique machine learning model based on their cognitive activity. Raw fNIRS data in the form of light intensity values was sent from the Imagent to a custom analysis system in MATLAB. During the music training task, the system calculated a time series of change in light intensity compared to a baseline average of that participant for each of the sixteen channels. For each music piece, markers were sent to the system using a python socket to denote when the task started and finished as well as what difficulty level the piece was. At the end of the trial, the system calculated the mean and linear-regression slope for each of the channels, resulting in 32 features (16 channels x 2 descriptive features) for each trial. It then fed these example trials to LIBSVM, a support vector machine classification tool. LIBSVM takes in a set of labeled training examples and creates a function using all of the features which can be used to predict which set a new examples belongs to. Parameters were calculated for each individual in order to optimize the model for each participant.

Learning Task Real Time Classification

While participants were learning, the machine learning model was making predictions of their cognitive state. For each level of difficulty in the BCI condition, a threshold was calculated based on data collected over a certain period of time. This threshold is what determined whether a participant was cognitively prepared to move on to the next level of difficulty. The system analyzed the last 30 seconds of real-time fNIRS data in order

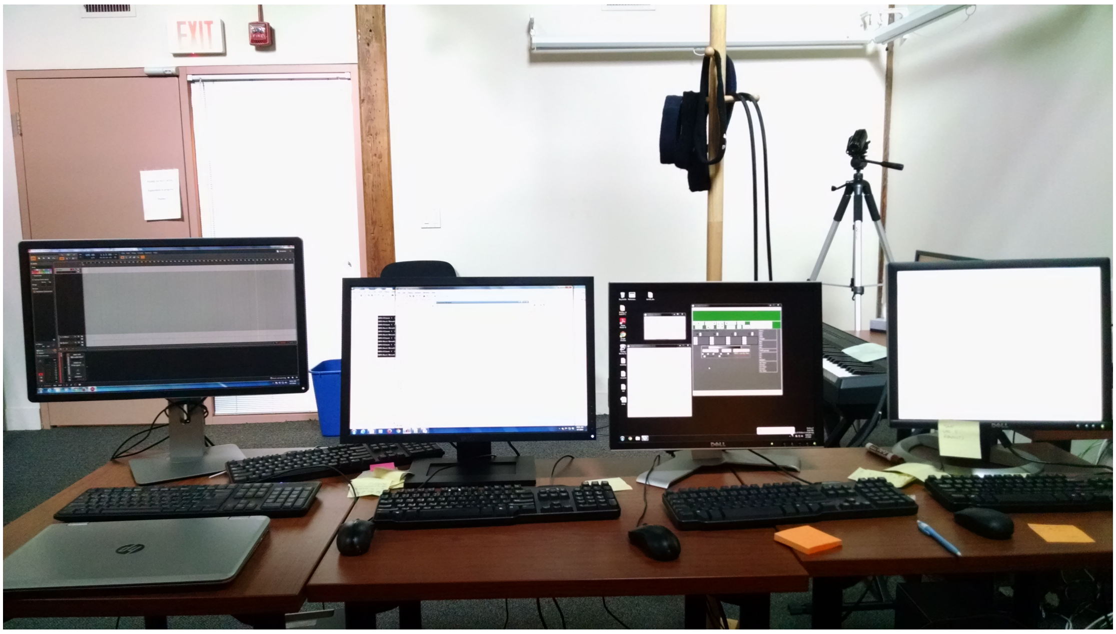


Figure 3.3: Lab computer setup from left to right: Bitwig MIDI acquisition, Matlab and LIBSVM processing, fNIRS raw input to Imagent, Java experimental code

to calculate a prediction and confidence interval based on the model that was created in LIBSVM during the training task. After enough predictions were gathered, the threshold was set at the 75th percentile of confidence values for both high and low cognitive workload classifications. The system was making predictions of the participant's current cognitive state. Then, once the threshold was set, the system was waiting for 65% of the last 20 predictions to drop below the threshold. These parameters and design are discussed below in section 3.6.1.

3.5.3 MIDI Keyboard and Bitwig

Participants completed tasks on a full-sized Yamaha keyboard with weighted keys. The keyboard was transmitting MIDI data via USB to a computer running Bitwig studio, a DAW that allows for MIDI recording and playback as well as visual displays of recordings. Participants' musical data was recorded with this equipment.

Figure 3.4: A participant sitting at the keyboard wearing the fNIRS device and reading music

3.6 Design Iterations and Pilot Studies

Prior to finalizing the parameters for the BCI and the music used in the learning task, several interviews and pilot studies were conducted. These pilots allowed us to explore the learning limitations of pianists, the effectiveness of our interface, and the feasibility of our design. From this preliminary feedback and data we were able to successfully create a BCI to assist pianists in learning music.

3.6.1 Pilot Studies

One of the first things that we needed to determine was what length of piece a pianist could feasibly learn in a short period of time. We tested a variety of pieces with varying lengths and difficulties. We also ran pilots giving participants different amounts of time to learn the music. Based on preliminary feedback and data from our first round of pilot studies, we determined that 4 measures of music with a mixture of quarter and eighth notes was a reasonable amount of music to learn in 8 minutes. We ran a couple more pilots with these parameters and uncovered an issue of varying skill levels of participants.

It became clear that some participants were much more skilled and had no trouble completing what was, for them, a relatively simple task. However, many of our beginner pianists struggled to learn much of the music at all with these parameters. Also, in the experimental condition we were not seeing much difference with intermediate/advanced pianists but a small effect was present for beginners. Based on this we determined that our system was more effective for those who were not very experienced on the piano. This is likely because experienced players are very comfortable in their abilities to learn music and have developed their own heuristics to do so, making our system irrelevant or even obstructive to the way that they learn. This observation was supported by the feedback of our pilot participants, discussed below.

Therefore, in our next iteration we eliminated eighth notes from the score so that it consisted of only quarter notes. We also extended the time to 15 minutes for each piece. These changes were made to accommodate beginner pianists. We started screening participants with a questionnaire (seen in section C) in order to determine if they were too skilled for our study. We ran a few pilots with this design but then discovered an issue with our BCI parameters.

Originally, the machine that made predictions based on brain activity was using an average value of the last 10 seconds or so of activity. Then, based on this value, the machine would make the switch if the average was above the threshold. However, we found that the switch was often happening much too soon or too late. We also noticed that the threshold for a specific level was sometimes being calculated when the participant wasn't playing all of the notes present. This presented problems because a) we wanted the threshold to be determined when a participant was playing or attempting to play all of the notes at once and b) an average was not a good gauge of brain activity over such a short period of time. We remedied these issues by doing the following. First, we altered the system so that it was looking for 65% of the last 20 predictions to be above the threshold in order to change. This parameter was narrowed in on through changes in multiple pilots. Then, we implemented a trigger activated by the experimenter that signalled when the participant had played at least a measure including all of the notes present so that the threshold calculation was more appropriate for the given level.

With these parameters for the music stimuli and BCI we were finally confident in a meaningful and effective system to be used for the final experiment. The data collected

collected for this paper was made with this final iteration of the experiment.

3.6.2 Interviews

Through our pilots and experimental trials we gathered a large amount of vocal feedback that led to some of the changes that we made and also confirmed our final decisions. In order to accommodate the skill level of beginner pianists, the Bach chorales did have to be altered which meant that the original counterpoint composition was no longer intact. Because of this we were concerned that the pieces may no longer have been comparable in difficulty. We asked pilot participants what they thought of the difficulty of the pieces compared to each other. A few participants felt that the piece used in the BCI condition was objectively easier so we asked them to elaborate and showed them the pieces side-by-side. The following is a participant's response when asked to explain the differences in difficulty while examining the scores:

> *"I felt the second one [BCI] was easier...I was going to say that the second one had more accidentals, but I guess not...Here it looks like there are bigger gaps between ntoes...Maybe I'm just thinking that it [BCI] was easier, they're actually very similar."*

This was a common reaction when we asked people to explain the differences side-by-side. Some people felt as though the BCI condition was an easier piece to play but then could not justify why. The response above also comes from someone whose primary instrument has been piano for the last 13 years and is majoring in music at school. This is a good indication that our adjustments and alterations to the music did not compromise the comparable difficulty of the pieces.

When determining our specific parameters, we actually ran participants with multiple parameters in our pilots. We got feedback from participants about the timings and eventually set our final parameters. The following is a response from a participant who ran through the BCI condition with several different threshold parameters, with the last time being our final parameters. This participant was asked to comment on the timings of the changes.

> *"The last time [with final parameters] was like perfect timing where I was getting really comfortable with it and then it would get more difficult so it kept me on my toes. The last time was really good timings.*

Responses and reactions like this are what led us to our final parameters and gave justification for the threshold determination process as well as how we triggered the BCI based on this threshold. The above response came from a beginner pianist with only 2 months of experience which was a good representation of the skill level that we targeted in our study.

In addition to finalizing the parameters through pilots, we also observed that participants needed time before they could play each level with both hands. Some participants also commented on this such as:

> *"Sometimes I would switch back and forth between right and left and just practice one line at a time. I was more comfortable with the right hand by the time that it changed."*

From these comments and further discussion we realized that our system was correct in determining when to change levels but was unaware that the user had only learned that level with one hand, hence it was switching between difficulty levels too quickly. In order to make sure the system was responding to the cognitive load of a participant when playing all of the notes we determined that a trigger was necessary to indicate when a participant had been playing all of the notes in the music for at least a measure. The experimenter sent a message to the system by means of key press when the participant had played all notes of each level for one measure. This can be implemented with automation but as it was not within the scope of this study, was implemented in Wizard-of-Oz fashion.

Chapter 4

Musical Data Analysis

4.1 Overview

In this experiment there is a variety of software used during trials as well as for data collection and analysis. When the experiment was being run, MATLAB and Java code was used to process and utilize the fNIRS signals in order to control the adaptive condition of the experiment. Music software along with Python code was then used to record and analyze participants performances. I will be focusing on the Music software and Python code as it was one of my main contributions to the design and execution of the study. I will also outline the manual assessment method that was used to measure accuracy in performance.

One of the biggest challenges in this study was determining the best way to assess how well a piece of music has been learned. There is score-following software that exists that tries to take on this task, but we found that no existing method provided us with all of the information that we were looking for so a decision was made to create our own method of performance analysis based on other research in this area.

4.2 Musical Assesment

Performance assessment of many kinds can be found in a wide range of research. For this study, we were faced with the task of assessing musical performance on the piano. Standards for assessing musical performance are not well established and are often done through judgements based on a certain criteria [24]. However, it has been shown that there are problems with assessing the quality of musical performance through judgement based systems [25]. For the purposes of this experiment we needed a reliable measure of how well someone learned a piece of music. Instead of relying on subjective judgements, we decided to focus on quantifiable data the notes themselves. This led to the research of score following software as an evaluative measure of performance.

Score following is, in itself, a very dense and widely studied field. Only in the past decade or so have there been dependable programs that follow along as people play piano. The difficulty with score following is creating a reliable algorithm that handles the non-deterministic nature of musical performance. The software for score following applications must make predictions about where a person currently is in the music and what they will do next. It also has to take errors and mistakes into account which becomes very difficult to program. When designing this study we considered the possibility of creating our own, basic score follower but quickly found through research that it would be outside the scope of this project. From this research we gathered information on the different ways that score following software assesses performance.

As mentioned above, there are no standardized ways to assess musical performance so we looked to score following for information on quantifying a users performance. In our research we found that score followers classify events (a user playing a note) in a number of ways. There are incorrect events, extra events, and correct events. Most simply, correct events occur when a user hits a note at the correct time. An incorrect event occurs when a user hits the wrong note while also playing correct notes. Extra events occur when notes are played unnecessarily such as repeating notes to correct oneself or mistakenly hitting the keyboard [26]. With these parameters in mind we were able to devise a system for measuring piano accuracy.

One factor that consistently came up in our research was tempo. Many score followers use a fixed tempo to better follow the users because this makes each note time dependent. For our experiment we didnt want to force any participants to adhere to a specific time because this might affect the accuracy. Since we were measuring how well someone learned a piece of music and not how well they could follow our parameters we allowed them to play at their own pace. We didnt want to ignore tempo completely so timing data was recorded of their performance allowing us to measure their average tempo and how consistently they played at this tempo.

One last factor of assessment that we used was dynamics which was measured by looking at the average velocities (how hard or soft a participant played) for different sections of the music. This was just an additional measure to see how closely the participants were following and learning the piece. By combining accuracy, tempo, and dynamics measurements we could confidently create a profile of the users performance with objective data that could be quantified and analyzed for significant results. This was accomplished using a variety of different software tools described below.

4.3 Assessment Software

4.3.1 Python Library

In order to gather and analyze accurate MIDI data from participants, an external Python library called Pygame was used. This library provided a simple way to select and iterate through different MIDI inputs and outputs, parse raw MIDI data, and manipulate this data quickly and easily. The midi module of the library was used in some of the accuracy and tempo measurements. This library makes use of the MIDI ports of the system computer that the code is being run on and parses data using polling methods to gather the data as it is passed through the MIDI ports. It is simple to specify other parameters such as which MIDI port to poll, how long or short the buffer should be, and how many events to read in from the buffer at any one time. Once a MIDI event is read, the library makes it easy to access all data in the MIDI specification.

4.3.2 Bitwig Studio

In order to gather participant data in a graphical representation and record their performance, a MIDI sequencer was used. After researching various options from various developers, we decided on Bitwig studio. This software is very robust and has many features that are outside the scope of this experiment but included some great tools for analysis.

The first positive feature of this application is the interface. Not only is it a clean design but it is also very easy and intuitive to use. Due to the characteristics of the interface, all of the tasks that we performed using this software were simple and straightforward. The other useful thing about the interface was the clear graphical representation of participants performances. This graphical representation allowed for easy, precise accuracy measurements. The interface also made it easy to mark, in real time, specific points of a participants performance to easily recognize important areas later during analysis. Another useful feature of this program is the option and simplicity of routing an instrument to a MIDI output. Since some of the analysis was done using a Python script, routing the participants performance to a MIDI output meant that the Python script did not have to run in real time as we could essentially recreate the performance using Bitwig and a virtual MIDI driver (described below). One last useful feature of this software was the velocity measurement features. This function of the software allowed us to look at groupings of notes to examine the mean and distribution of the velocities in order to assess how well participants followed the dynamics of the piece.

4.3.3 LoopBe1 Virtual MIDI Driver

In order to analyze performances after the fact, a virtual MIDI driver was used called LoopBe1. This driver allowed us to send MIDI output from Bitwig (described above) and receive input in the Python program. The program needs a MIDI input in order to analyze the data. In order to avoid using unnecessary MIDI instruments to route the data, this driver was installed. Essentially, the data it receives from Bitwig is the same as a participant playing the piano. The driver then does the same thing the keyboard normally does by sending the data somewhere else. The difference is that this all happens internally so that the data can be played and processed in the same place. Another benefit to this driver is that all analysis can be done offline and after the fact. So if a parameter in the code needs to be changed, the recorded data can be run through the program as many times as needed.

4.3.4 Design and Implementation

All of the code and software mentioned above was used together to create a profile of each participant's performance to assess how effectively they learned the given piece of music. First, Bitwig is used to record the MIDI data in real time. With this data, MIDI is sent to a python program using LoopBe1 which then analyzes the timings and makes a count of each note played.

The Python code (seen in section A.1) gathers raw MIDI data using the pygame module. For each byte of data the program records what type of MIDI message it is, what the MIDI note is, and the timestamp of the data. When a message comes in, the program checks to see if it is a 'note-on' or 'note-off' message. If it's a note-on message then the timestamp is stored in an array. However, if the timestamp is within 150ms of the previous one that was recorded it is ignored because this means that the notes were essentially played simultaneously. This timing threshold was determined using data from pilot studies by manipulating this value to find an appropriate time frame. The timestamp is only needed for each beat that a participant plays, not every single note. The note-off timestamps and MIDI note numbers are also stored.

Once all of this information has been recorded and stored, the program then makes a few calculations. First, by using the total time the person took to play the piece and the number of beats that they played, it calculates the tempo as if the person had been playing with perfectly consistent gaps between each beat. It then looks at the amount of time between each beat and compares it to what the time gap should be based on the calculated tempo. This gives a way to evaluate how consistently people played to a certain tempo. A separate program (seen in section A.2) was run on the timing data in order to calculate the variance of each participant's tempo by condition. This code parsed a text file with the timing data and output a CSV file to be used in statistical analysis software. These calculations can be seen in detail in the code included in the appendix.

Once these calculations have been made, a file is produced including the following:

- a list of all notes played
- a list of all of the timings for each beat
- the number of beats played, the total time the performance took
- the average gap between notes
- the average tempo (in BPM)
- the range of gaps between notes (minimum and maximum gaps)
- the average variance of the note gaps
- the number of each note played (i.e. a list of how many times a given note was played)
- the total number of notes played

Some of this data is not used in this paper but is all likely to be used in future analysis of this experiment. In addition to this data gathered through python MIDI analysis, we needed some way to assess accuracy. As mentioned above, research led us to the conclusion that a score-following algorithm would not only be difficult to create but also not the best measure of performance without a very sophisticated algorithm that would be outside the scope of this program to create. So in order to make these assessments a manual analysis method was used in combination with the data gathered using the python code.

4.4 Manual Assessment Methods

Bitwig Studio provides a useful piano roll display of notes that have been played. Instead of listening to the performances to determine how well participants performed, the graphic representation was used and compared against a perfectly accurate performance generated using the MIDI file from the notation of each piece. With the visual representations side by side, I was able to analyze the performance in terms of errors and accuracy. Figure 4.1 shows an example of the layout for manual assessment. I analyzed each performance by hand with the aid of video and audio data.

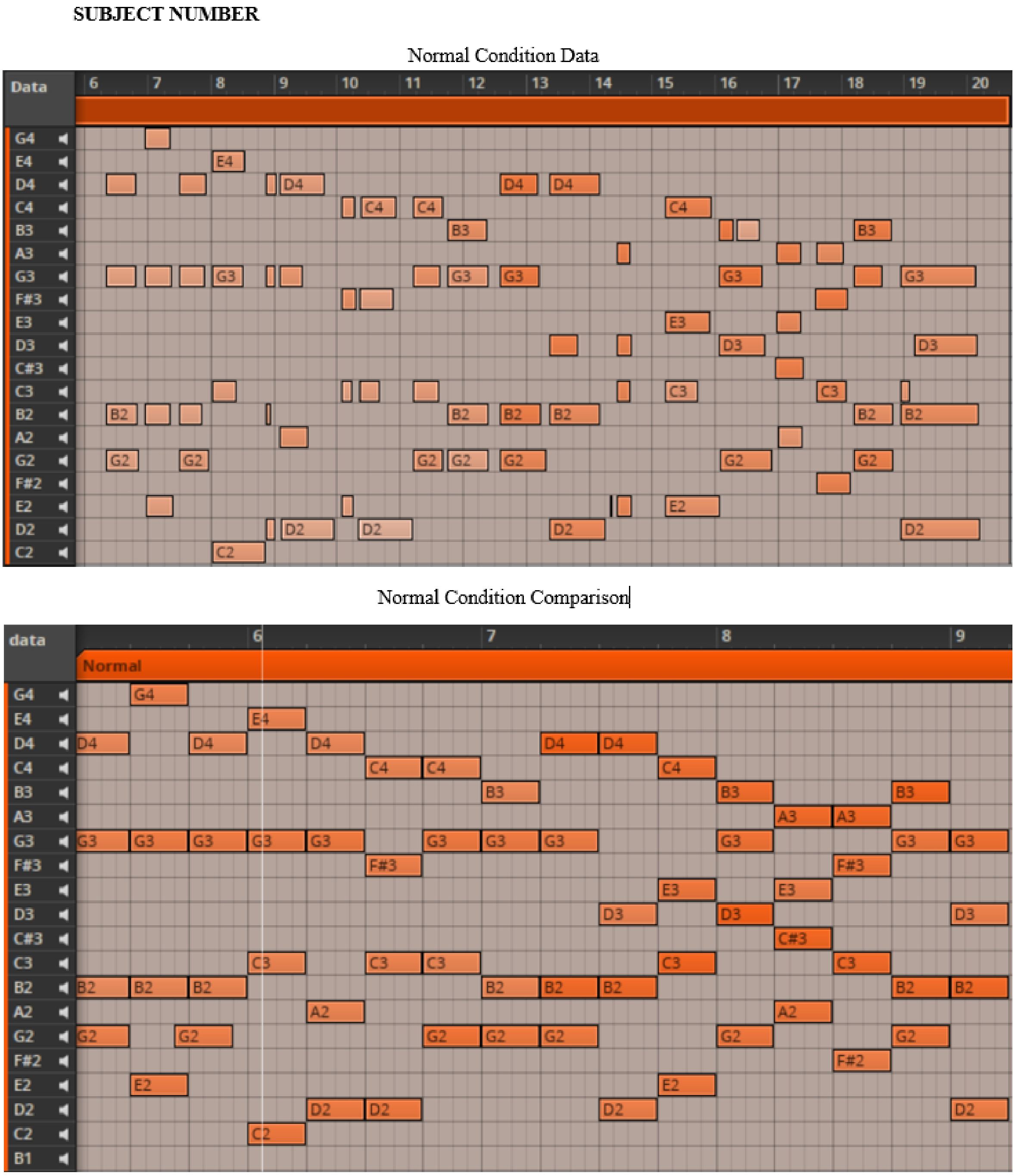


FIGURE 4.1: Visual comparison between user generated data (top) and data generated from MIDI files of the original score (bottom)

There were three events that were marked on these graphical displays:

- Correct notes
- Incorrect notes
- Extra notes

As a musician and pianist I was able to make distinctions between these notes when analyzing the performances. Having played piano for 16 years, I have a good understanding of how people play and make mistakes when performing a piece. Having also taught piano, I am familiar with common ways that pianists correct themselves while playing something that they are unfamiliar with. Using this visual representation in conjunction with video and audio data I was able to determine where mistakes were made and what types of mistakes they were.

A common mistake that many participants made was playing a correct note and then repeating the entire chord immediately while simultaneously trying to correct their original mistake. This can be difficult to assess in the visual representation because a determination needs to be made about which notes are repeats with corrections and which ones are unnecessary repeats. This is where video comes in handy to make these determinations. Another issue that comes up is when participants start over from the beginning after playing a few notes. In these cases I have to decide which notes to count as extra and which ones to mark in terms of correct or incorrect. Again this is determined on a case by case basis.

Once correct, incorrect, and extra notes have been marked on the graphical representation the next thing to be recorded is errors. An error is a temporal event that includes either an extra or incorrect note. However, only one error is recorded for any temporal grouping of notes that includes any number of incorrect or extra notes. This data parameter was included to avoid over-penalizing participants who tended to repeat notes frequently. So with this measure, 4 extra notes that occurred simultaneously were also recorded as one temporal error. In section D an annotated example of this manual assessment method is outlined.

4.5 Challenges

There were many challenges that I faced when designing data collection for this project. The biggest difficulties came in the first stages of the experiment when we were still considering the creation of a custom score-following software. It took a few iterations of code and music stimuli before this final method was settled on.

In designing a score-following software in the early stages of this project I came upon the issue of parsing a piano performance in comparison to another. This difficulty was largely due to the fact that measuring a piano performance needs a non-deterministic algorithm to assess accuracy and keep track of where a participant is in the score. One possibility that I explored was gathering temporally related notes into small data arrays and comparing all permutations of these notes to a pre-recorded data from the score generated MIDI file. This method was effective in measuring incorrect notes. However, it would only give an accurate measure if a user did not play any extra notes because then the comparison algorithm had different sized data arrays to analyze which was not a problem addressed in the code.

After exploring these score-following options and conducting more research, we decided that accuracy measurements could be more easily and reliably measured by hand. So this left timing and tempo data to be recorded programmatically. One issue faced when processing MIDI data for tempo measurements was determining the best way to not only measure an average tempo but also assess the consistency of a performance. At first, individual variations were averaged together but, after consulting with individuals more familiar with statistical analyses, this was determined to be a poor measure of consistency. The final method of consistency measurement was done by finding an individuals variance in gaps between notes (using the MIDI timing list produced) for each condition, and then comparing the variances from the two conditions against each other.

One last difficulty faced was consistency in the manual analysis. As I was the only one working on these analyses I had to stay as unbiased and consistent as possible. I was able to establish a consistent scoring process through analyzing all of the pilot study data. Then, once the experimental data was being collected I needed to make sure that my assessments were reliable. To do so, I would randomly mark performances that I had previous recorded without knowing what my original judgements were. These repetitions showed little to no variation from the original results that I had scored.

4.6 Future Designs for Musical Analysis

Working on this project, especially in musical analysis, has been very educational and encouraging for possible future work. As mentioned before, score following is a problem outside the scope of this project and could be a separate thesis. After manipulating MIDI data for accuracy and tempo measurements, I think that spending more time creating MIDI parsing software could prove very useful.

In future versions of this experiment, I would improve both the Python code and the manual analysis. For the Python improvements, I would add in a confidence buffer to determine if notes are happening simultaneously or quickly after one another. This would improve the accuracy of the number of beats measurement. I would also calculate variance for each condition in the code so this statistical analysis would not have to be done after the fact. This would save a considerable amount of time. As for the manual analysis, I would develop some way to overlay the participant's performance with the score generated MIDI data instead of a side by side comparison in order to make these comparisons quicker. I would also develop some simple marking software to make these note judgements so that the values of correct notes, incorrect notes, extra notes, and errors didn't have to be tallied by hand.

Overall, I was very satisfied with the software and manual methods used for data acquisition and I am confident in the reliability and validity of the results.

Chapter 5

Results

5.1 Musical Analysis

Results from statistical analyses of the musical data showed multiple significant effects, suggesting that the BCI was successful in helping beginner pianists learn a piece of music. Prior to running any significance testing we used a Shapiro-Wilk test to check each condition for normal distributions. Both the BCI and normal condition needed to have a normal distribution in order to perform parametric testing. The results of this testing can be found in table 5.1. From these results we found that the only dependent variables with normal distributions in both conditions were total notes played and percentage of notes played correctly.

Once normality tests were run, we ran tests to look for statistical significance. Since this was a within-subjects design we used paired tests to look for an effect of condition (Normal or BCI) on each of our dependent measures. The means and standard errors of each measurement can be found in table 5.2. We ran a Wilcoxon Signed-rank test on all non-parametric data (not normally distributed) which included the number of correct notes, number of incorrect notes, number of extra notes, number of beats, number of errors, percentage of beats that include errors, total time played, mean gap between notes, and average BPM. Results of these tests can be found in table 5.3. For the normally distributed data, which included the total notes played and percentage of

	Normal		BCI	
Measurement	W	p	W	p
Number of correct notes	0.9208	0.09022	0.9029	0.03976
Number of incorrect notes	0.874	0.01135	0.8759	0.0123
Number of extra notes	0.7904	0.0004698	0.8105	0.0009553
Total notes played	0.9451	**0.2739**	0.9398	**0.2157**
Percentage of notes played correctly	0.9815	**0.9447**	0.9335	**0.1621**
Number of beats	0.8478	0.003892	0.872	0.01041
Number of errors	0.9239	**0.1038**	0.9089	0.05225
Percentage of beats that include errors	0.9653	**0.6283**	0.9016	0.03754
Total time played	0.7945	0.0005413	0.803	0.0007305
Mean gap between notes	0.8893	0.02178	0.7409	9.33×10^{-5}
Average BPM	0.9315	**0.1473**	0.9	0.03506
Tempo variance	0.7656	0.0001517	0.822	0.00113

Table 5.1: Results from Shapiro-Wilk tests for each parameter by condition. Normal distributions marked in bold.

notes played correctly, we ran paired t-tests. The results of these tests can be seen in table 5.4.

Using the music data from all trials, including the performances from the follow-up studies, we saw a significant effect of condition on the number of incorrect notes (W = 15.5, $p < 0.01$), the percentage of notes played correctly (t(20) = 2.6506, $p < 0.05$), the number of errors (W = 7, $p < 0.01$), the percentage of beats that include errors (W = 22, $p < 0.01$) and the total time played (W = 43, $p < 0.05$). These results suggest that the BCI condition was a better system of learning music which is consistent with our original hypothesis.

	Normal		BCI	
Measurement	Mean	Standard Error	Mean	Standard Error
Number of correct notes	44.238	2.697	46.667	3.235
Number of incorrect notes	7.476	1.251	4.809	1.039
Number of extra notes	17.524	3.818	14.667	3.435
Total notes played	69.238	6.022	66.143	5.701
Percentage of notes played correctly	67.595	3.188	74.052	3.663
Number of beats	21.667	1.119	20.667	1.031
Number of errors	11.095	1.421	8.238	1.318
Percentage of beats that include errors	48.576	4.353	37.509	5.157
Total time played	61.038	9.045	52.408	7.596
Mean gap between notes	2.727	0.302	2.492	0.328
Average BPM	27.619	2.865	30.429	2.645
Tempo variance	3052737.344	181704.489	2040288.312	104574.041

Table 5.2: Means and Standard Errors of measurements by condition

Looking at table 5.2, there are clear trends in favor of the BCI condition in some metrics that did not show significant results. We thought that this may be affected by

Measurement	W	Z	p	effect size
Number of correct notes	132.5	-1.481	0.1355	–
Number of incorrect notes	15.5	2.904	**0.004042**	0.448
Number of extra notes	76.5	0.679	0.4686	–
Number of beats	71.5	0.924	0.353	–
Number of errors	7	3.371	**0.00169**	0.521
Percentage of beats that include errors	22	3.249	**0.0005102**	0.501
Total time played	43	2.519	**0.01013**	0.389
Mean gap between notes	69	1.616	0.1111	–
Average BPM	128	-1.797	0.06673	–
Tempo variance	261	1.315	0.1982	–

TABLE 5.3: Results from Wilcoxon Signed-rank test. Significant results shown in bold.

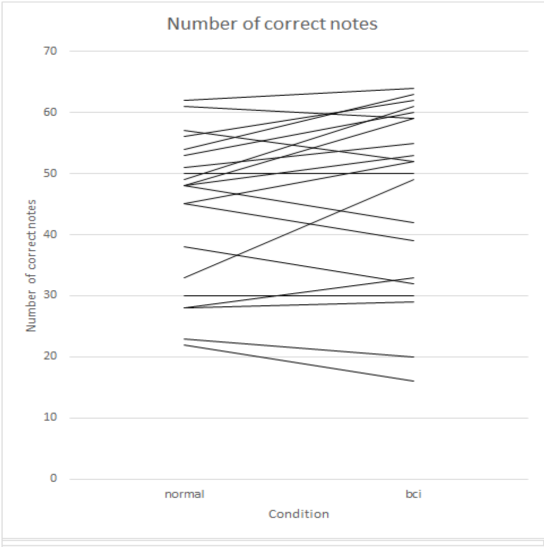
Measurement	t	df	p	mean of differences
Total notes played	-0.8393	20	0.4112	-3.095238
Percentage of notes played correctly	2.6506	20	**0.01535**	6.457143

TABLE 5.4: Results from paired t-tests. Significant results shown in bold.

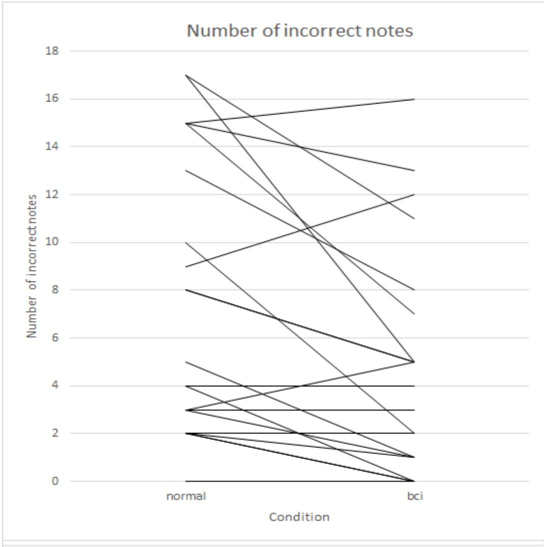
the fact we included data from the follow-up trials aimed to measure retention of the pieces. With this in mind, we also ran statistical tests on the data without the results from the follow up studies. From these analyses, in addition to seeing significant effects in the other measurements again, we also saw a significant effect of condition on the mean gap between notes (W = 18, $p < 0.05$) and average BPM (W = 63, $p < 0.05$) (table 5.5). This suggests that while participants seem to be learning the pieces better in the BCI condition, this effect may not be as prevalent during follow up trials for all of the measurements. A visual representation of all results can be found in figures 5.1 and 5.2.

Measurement	W	p
Number of correct notes	78.5	0.1086
Number of incorrect notes	10	**0.0247**
Number of extra notes	39.5	0.7005
Number of beats	38.5	1
Number of errors	5	**0.008438**
Percentage of beats that include errors	1	**0.0002441**
Total time played	18	**0.02954**
Mean gap between notes	18	**0.02954**
Average BPM	63	**0.00862**

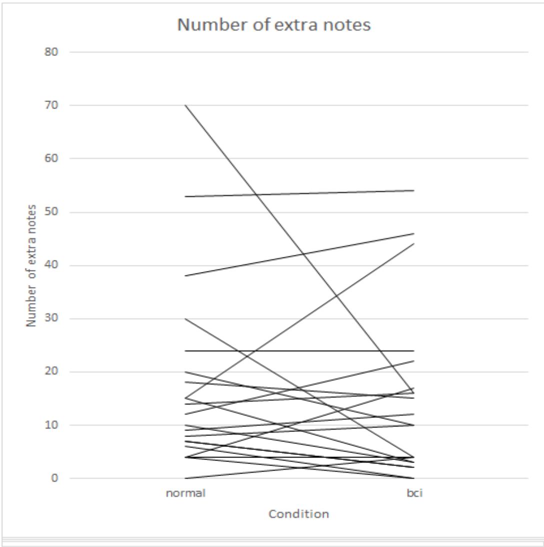
TABLE 5.5: Results from Wilcoxon Signed-rank test when data from follow-up trials are excluded. Significant results shown in bold.



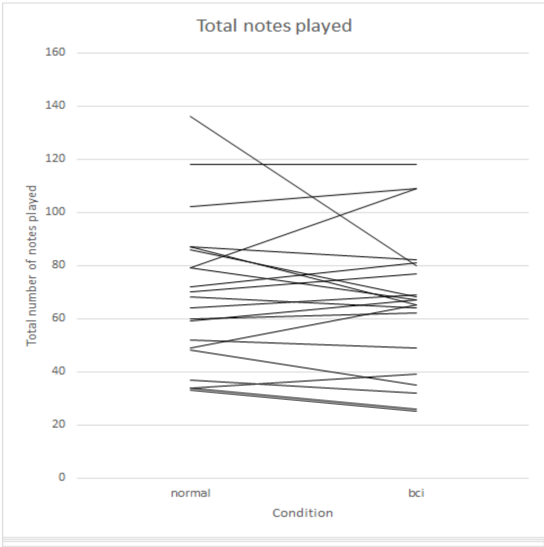
(A)



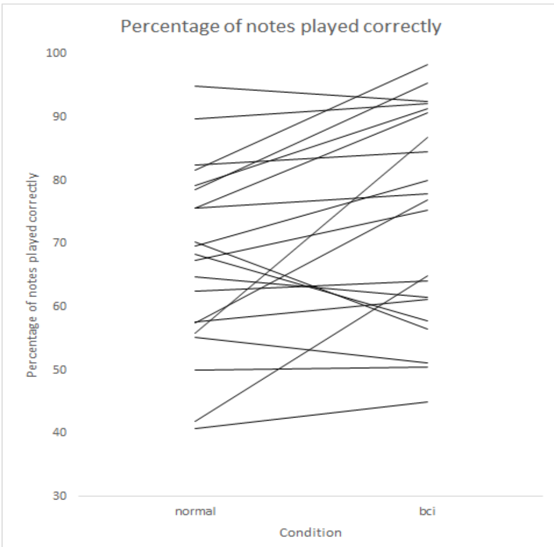
(B)



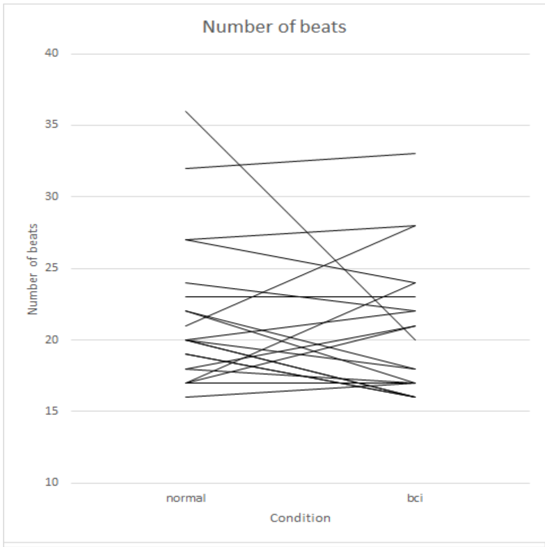
(C)



(D)

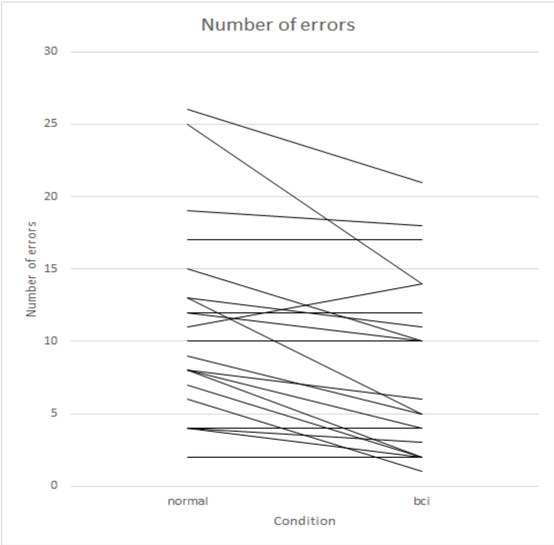


(E)

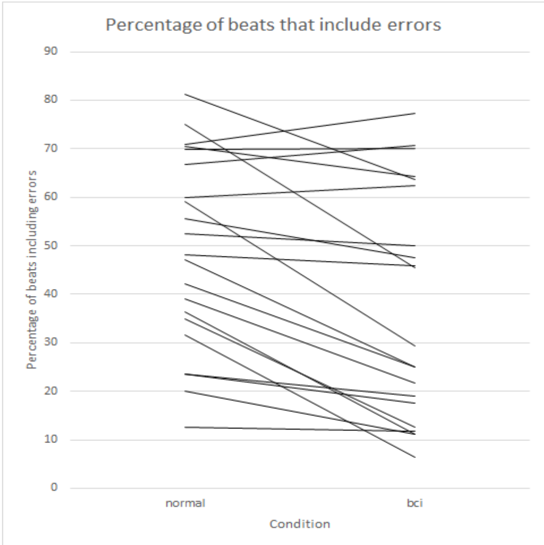


(F)

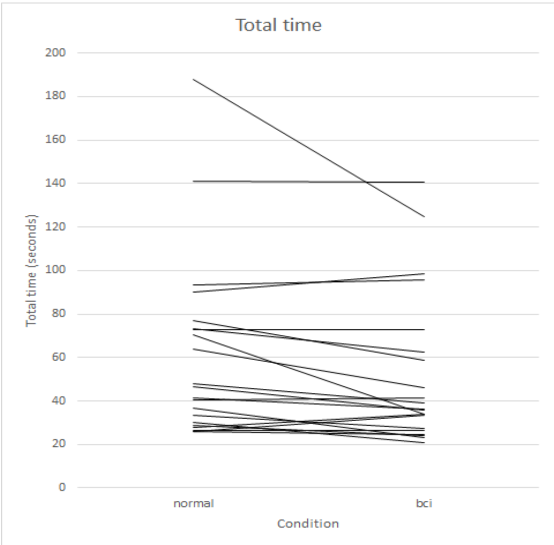
FIGURE 5.1: Results of musical data measures showing within group comparisons



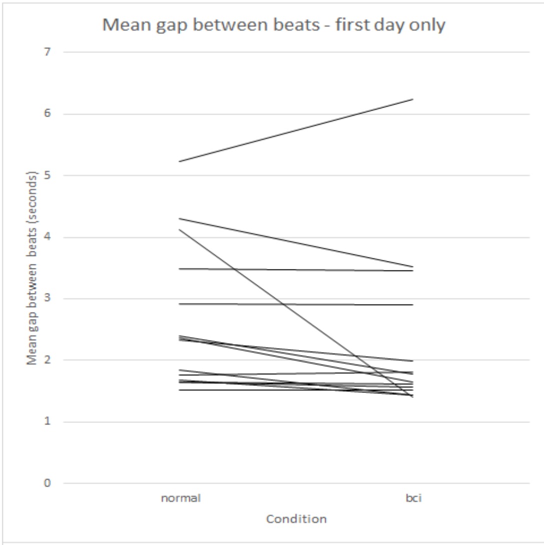
(A)



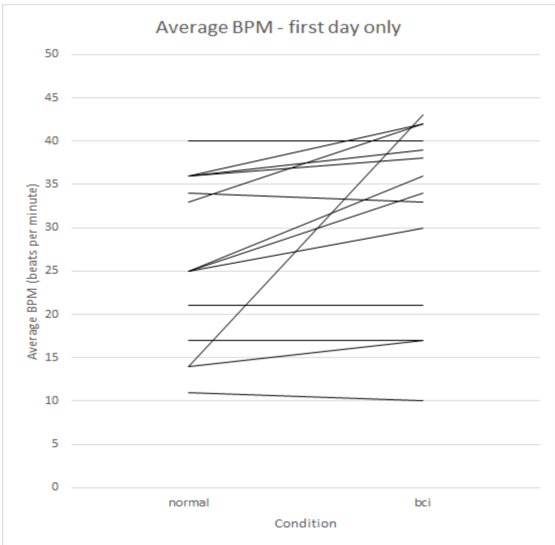
(B)



(C)



(D)



(E)

FIGURE 5.2: Continued results of musical data measures showing within group comparisons

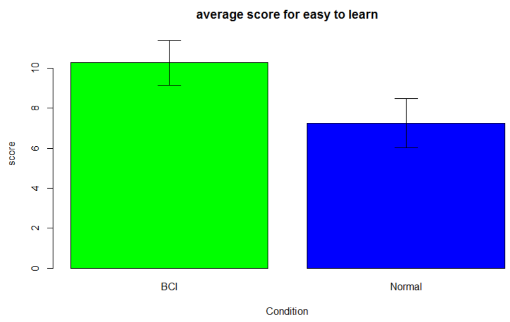
5.2 Questionnaire results

In addition to the musical data, we examined data gathered from questionnaires filled out by participants after completing the experiment. These questionnaires can be seen in full in section C. Responses were recorded on a Likert scale with statements that participants rated from 0 to 20, with 0 meaning that statement was not true for the participant at all and 20 meaning it was very true. This data is all self-reported but shows noticeable differences between conditions. Mean comparisons can be seen in figure 5.3 with noticeable differences in how easy participants felt it was to learn (figure 5.3a), how well they felt they mastered a piece (figure 5.3c), and how correctly they felt they played the piece (figure 5.3d). While they weren't all signifivant, all questions showed responses favoring the BCI condition which is consistent with our hypothesis that the BCI helped people learn better.

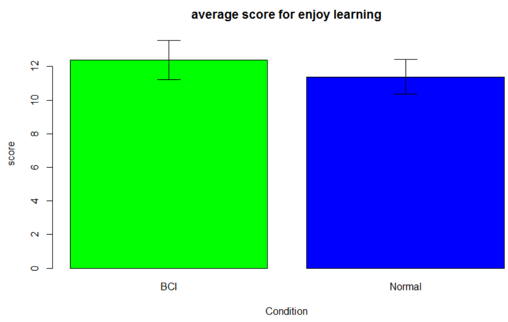
In addition to filling out the questionnaire about each of the pieces that they learn, participants were asked to choose their favorite of the two systems. Out of our 14 participants, 9 preferred the BCI system and 5 preferred the normal system of learning music. We thought that these responses would show a higher number of individuals who preferred the BCI condition based on the musical data and questionnaire results. However, table 5.6 shows that among those who preferred the normal condition, they performed better in the BCI condition in all dependent measures. To investigate this result and the effectiveness of our BCI as a learning tool, we looked to participant feedback in interviews.

	Normal		BCI	
Measurement	Mean	Standard Error	Mean	Standard Error
Number of correct notes	43.375	2.735	44.625	3.546
Number of incorrect notes	5.625	1.056	4.875	1.296
Number of extra notes	19.25	4.979	6.875	1.849
Total notes played	68.25	7.395	56.375	4.901
Percentage of notes played correctly	69.263	3.762	82	3.507
Number of beats	22	1.335	17.75	0.477
Number of errors	10.25	1.673	5.875	1.337
Percentage of beats that include errors	42.625	4.815	30.425	6.071
Total time played	71.524	13.323	56.869	10.586
Mean gap between notes	2.923	0.384	2.989	0.472
Average BPM	26.75	2.827	29.125	3.298

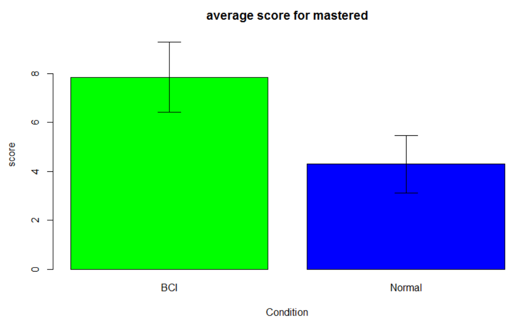
Table 5.6: Means and Standard Errors of measurements by condition of participants who rated the normal condition as their favorite system of learning



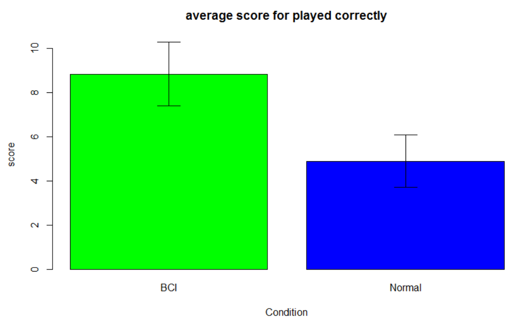
(A)



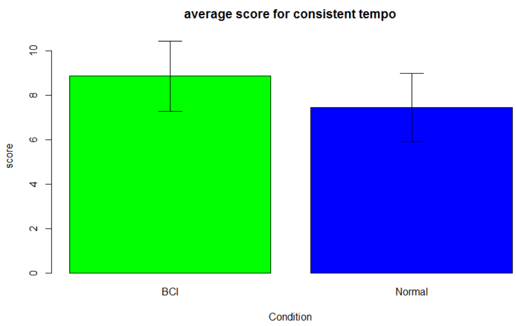
(B)



(C)



(D)



(E)

FIGURE 5.3: Results of follow-up questionnaire responses for BCI (green) and normal (blue) conditions.

5.3 Interview Responses

Because our system was such a novel interface, we were interested in what it did effectively and what the design flaws were. We could not extract this information from any of our musical data or self-report responses so we relied on a fairly in depth follow up interview with users of the system. The main questions we were interested in were why participants preferred the system that they indicated and what they thought of the BCI condition and its parameters.

5.3.1 Timing Parameters of BCI condition

The goal of the BCI was to change the level of difficulty at a time when the participant's workload was low enough to handle more information and a higher level of difficulty. When participants were asked about the timings of these changes, which were controlled entirely by the system, their feedback was generally positive:

> *"I thought it was good timings because by the time I learned, it gave me enough time to learn the individual lines, one by one."*

> *"I thought they were good times for changes, all of them."*

> *"Having a timing system can be jarring, you should only add new things when you know that the person has completed the existing part, but these timings were fine."*

One participant even seemed convinced that the experimenters were triggering the changes:

> *"I wasn't sure if you were controlling it or not because when it was added was a pretty appropriate time for me to add on to a part. Especially because in the beginning, one line, for me at least, is very easy to sight read so just getting that melody in my head and figuring out fingering for that one line*

> *and then adding on to it very quickly afterwards was helpful. I felt the timing was pretty good. I wasn't sure if it was timed or if you were like, oh she's done with this part, so add on to the second part."*

From these results we were able to confirm our assumption that the parameters set based on numerous pilot studies were effective and supported our original hypothesis. Some participants were not completely satisfied with the timings but in cases like these comments were similar to phrases like this:

> *"Sometimes I wouldn't notice it would change until I would look at the screen, it was a little confusing when I would look up. Yeah, it changed when I had learnt pretty much what I could learn before it changed, it was enough time to learn it,"*

or

> *"I thought they [the timings] were pretty good I think it seemed pretty good overall, the only thing would be the first one was a lot easier to learn because there was only one line, but it wasn't that bad,"*

which are not completely negative. Overall, there were no standalone or overtly negative comments about the timings of when the BCI system determined it was appropriate to increase difficulty. It is clear that participants felt like the changes were happening at an appropriate time which implies that we able to successfully measure cognitive load and manage it by altering the difficulty of a given stimulus.

5.3.2 BCI Feedback

The majority of participants preferred the BCI condition to the normal condition and participants on average felt like they had learned and played the BCI piece better (figure 5.3) so we asked them what they thought about it. Many participants felt that breaking up the music was a helpful method of learning and gave them a better understanding of the piece:

> *"It seemed easier to learn and even thought the score changed it made me learn piece by piece, it progressed from more simple to more complex so I liked that."*

> *"I felt like I was able to better understand what each part was doing, instead of trying to take it by chunks. the way the piece was presented made it seem easier, because I was able to break it up maybe that made it easier for me, even though it wasn't actually easier."*

> *"It was easier to learn because it came in parts they gave me one line and then they added another note on each hand so that was easier instead of just having to do it all at once. And I think that way I caught the melody more so it was easier to remember that."*

For most individuals, this progressive method of learning was something that they both enjoyed and was beneficial to their learning and understanding of the piece. What is most interesting however is not the comments from those who preferred the BCI system, but rather the feedback from participants who preferred the normal condition. In many cases, even though they didn't necessarily like the BCI system, they still acknowledged that it had beneficial effects on their learning. The following comments come from individuals who preferred the normal condition:

> *"The first piece was easier for me but it was hard learning it but playing it was easier. And I don't think I made as many mistakes."*

> *"The second one [BCI] shows up little by little and I practice little by little so it helped me to learn better. I think it's a better way to learn but it makes me feel there's too much coming up."*

Similar remarks were also made in the follow up studies:

> *"I thought it was interesting that the first piece I played today [normal] even though I thought it was easier yesterday, I felt that playing it again today I*

> *didn't remember it as well and I was basically sightreading it again and didn't really remember the fingering. Whereas the second one [BCI] since I played it so many more times, broken up and together, it was still in the muscle memory in my fingers and I was able to play it better, I didn't stumble as much."*

So even though some participants preferred the normal condition, the BCI system was still a useful tool for learning music, often better than the normal condition. These responses are consistent with the results in section 5.2 where those who preferred the normal condition actually performed better.

5.3.3 Normal Condition Feedback

One of the most prevalent responses we got from participants about the normal condition was that it was in some way overwhelming to have so much to learn at once:

> *"The first time [normal] it was kind of overwhelming to have all of the things at the same time, so much so that I just ignored the bass part and only did the right hand. I can't really remember the first piece. I kinda gave up on one of the parts."*

> *"The first way [normal] I tried to learn it measure by measure, which was more daunting then trying to do a complete melodic thought and then add on top of that."*

> *"It just surprised me that overall there were the same number of notes to learn and it wasn't like each hand had to handle more notes but it just felt a lot harder to learn because it was coming all at once and it was alittle more difficult to grasp what it was supposed to sound like and the underlying melody."*

> *"The music came all at once, and I kind of got scared."*

These responses of the normal condition being "overwhelming" is consistent with our findings that participants did not perform as well in our objective measurements of piano performance.

5.3.4 Additional Feedback

In addition to all of the positive feedback to the BCI condition, there were some recurring criticisms of the system. The most prevalent criticism we received was about fingering positions on keys when trying to learn with the BCI system:

> *"I didn't like the changing lines because I'd change the fingering, like I would learn it one way and then it would switch it and I would have to change the fingering."*

> *"I didn't feel like I could confidently figure out the fingering for the parts, In terms of actually mechanically playing the piece it was more difficult."*

Some of the participants felt that they had to re-learn where to put their fingers when new voices were added to the music. This was a concern that we had early on in the design but decided not to give fingering suggestions for each line. While giving fingering positions would prevent users from needing to reposition their hands, we felt that it was too much instruction in how to learn the piece. Our whole system is based on the idea that you are still learning at your own pace with whatever methods you normally would use. Part of learning a piece is figuring out appropriate positions of your fingers. Even when learning a piece normally pianists tend to change fingerings until they find out what works best. However, this would still be something that could be improved on in future iterations.

5.4 fNIRS Results

In addition to showing improved performing and learning on the piano, we also set out to show that our system is an effective measure of high and low cognitive load when playing piano. The BCI condition made changes based on brain activity dropping below a certain threshold. Our system made predictions as to when this threshold was reached based on the model created for each individual in the training task. In order to obtain this data we ran the mean and slope of fNIRS data for each individual into LIBSVM to create a profile of all of our participants. Figures 5.4 and 5.5 show the data collected and averaged across all participants from the training task. Highlighted graphs in these figures show the most significant channels; longest distance (3cm) from the sensor. Figure 5.6 shows an average of brain activity in the training task for all participants separated by hemisphere.

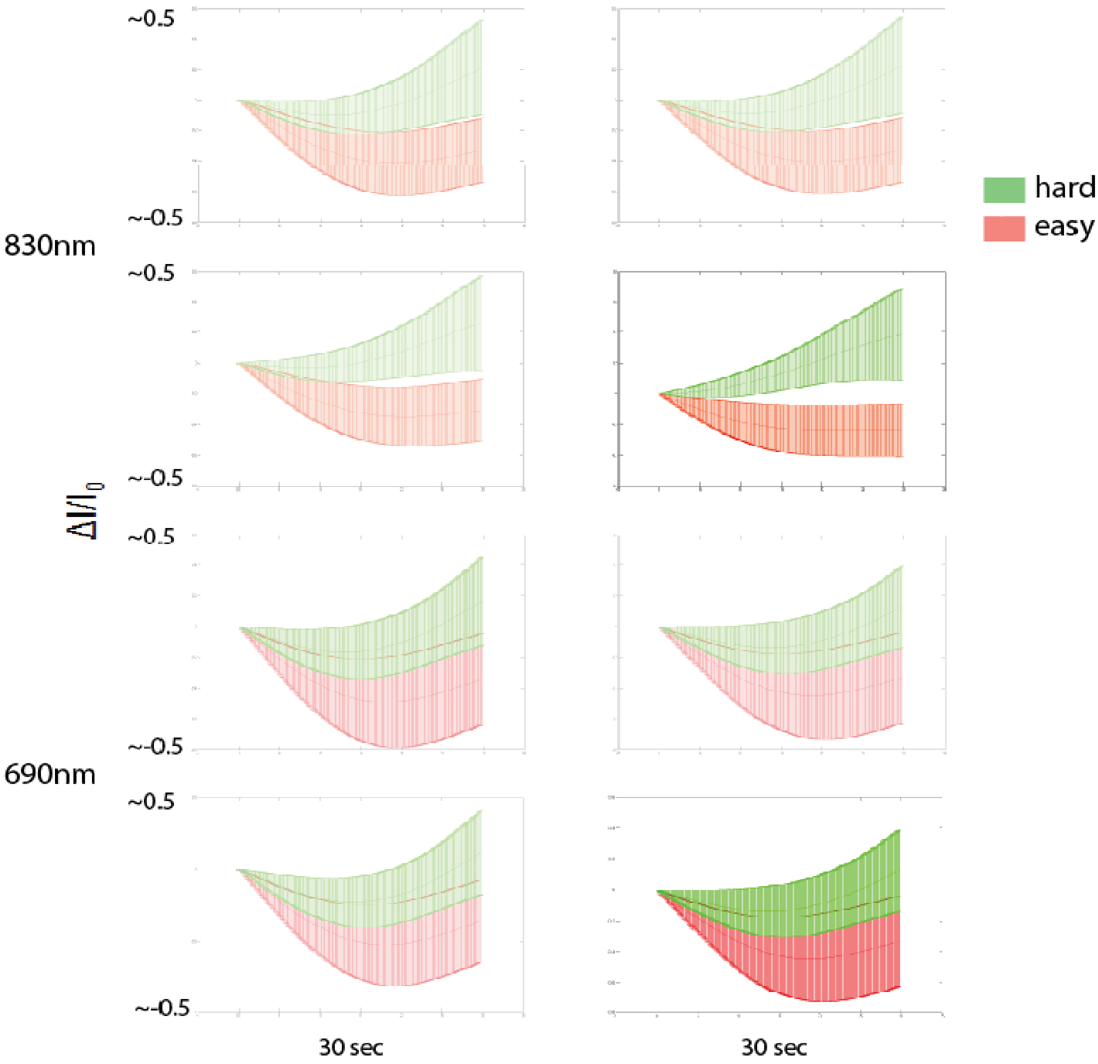


FIGURE 5.4: First 8 channels of averaged fNIRS measurements showing relative change in optical intensity of all 14 participants across all trials of training task during the easy (red) and hard (green) pieces

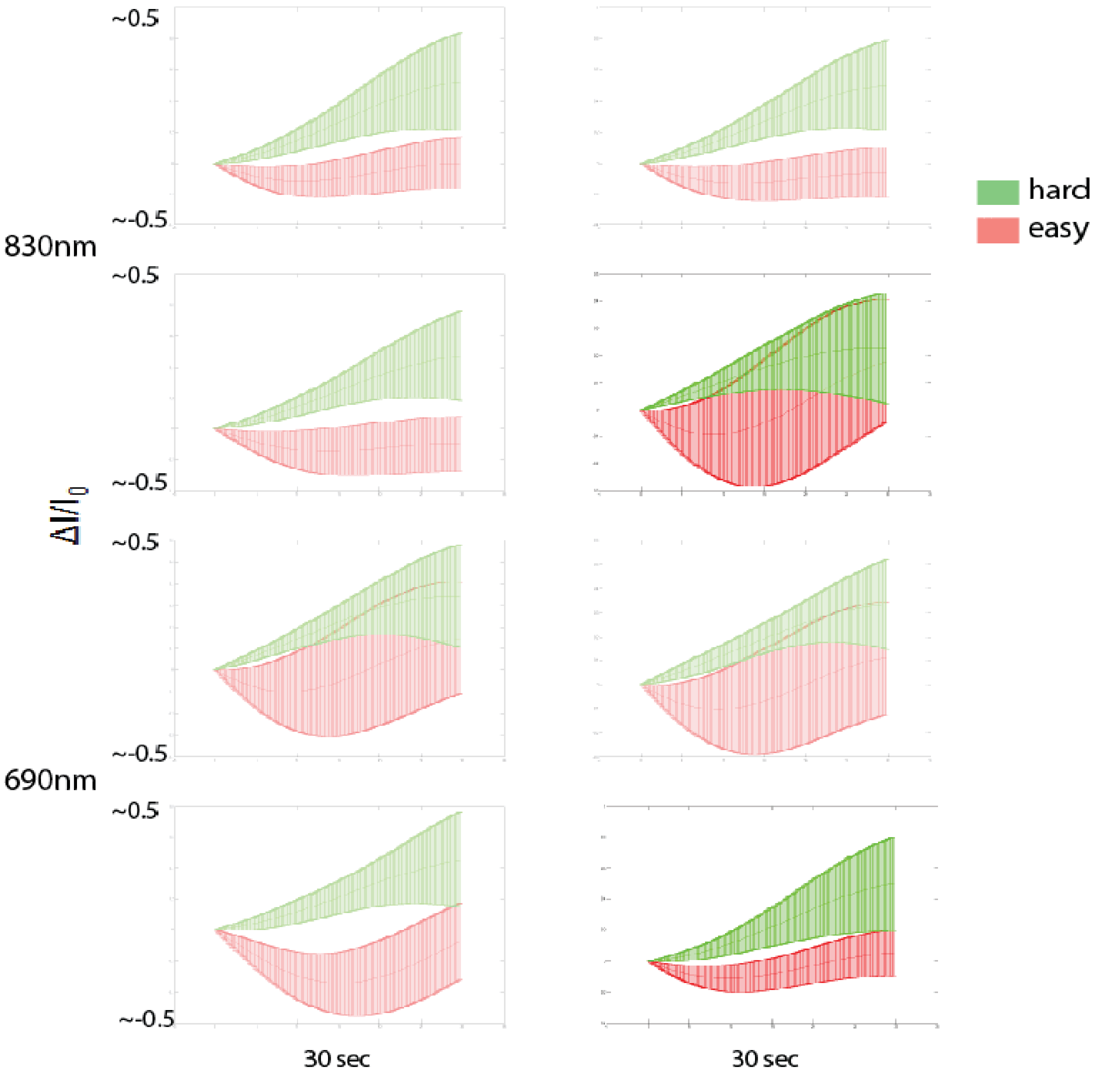


FIGURE 5.5: Second 8 channels of averaged fNIRS measurements showing relative change in optical intensity of all 14 participants across all trials of training task during the easy (red) and hard (green) pieces

From these graphs we can see that there is a clear distinction between brain activity in participants when playing difficult versus easy pieces. With this result we can be confident that our models for cognitive load that were controlling the BCI condition were making fairly accurate predictions as to the current cognitive state of the user.

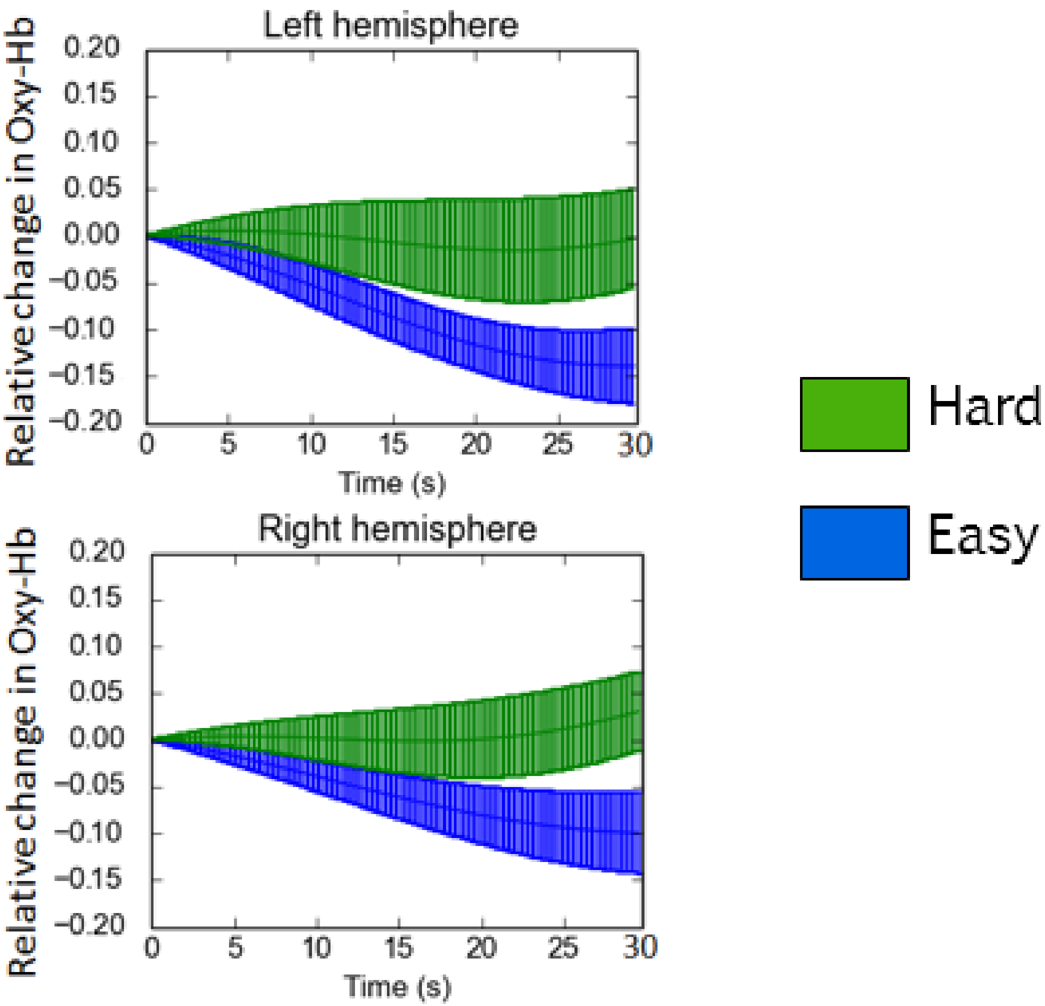


FIGURE 5.6: Average of relative change in Oxy-Hb in all participants. Showing difference in brain activity in each hemisphere during training task while playing easy (blue) and hard (green) pieces

Chapter 6

Discussion

We believe that we have successfully created and evaluated a real-time musical BCI which adapts to a user's cognitive state and improves the learning of a piece of music leading to increased performance accuracy and speed. We also believe that this is the earliest example of a successful BCI that adapts to cognitive load in order to improve learning. This is shown through our musical data measures, questionnaire results, and interview responses. We argue that this system of learning is an improvement to the traditional way of learning music.

In the musical data collected we saw that performances were better in both accuracy, mistakes made, and tempo. The fact that all of these measures showed improvement is a strong indication that the studied method of learning was helpful in learning the piece better as a whole as opposed to simply improving speed. This can also be seen in the feedback from the questionnaires and interviews; participants rated the piece in the BCI condition easier to learn on average and also reported a better understanding of the music and its structure in some cases.

Not only was it an effective interface in improving learning but the majority of participants enjoyed learning more with the BCI and subjectively felt better about their performances after learning with the BCI. Even though this was self-reported data it is still significant to the effects of our system. Self-assessment has been shown to be a useful method in performance evaluation [27] so even as a subjective measure, these responses give more support to our original hypothesis.

6.1 Implications

Shown to be successful, our system could theoretically be used in the context of piano lessons in order for the student to have an improved learning experience and better understanding of the pieces. Of course, this is not feasible due to the accessibility and price of equipment similar to what was used in this experiment. More broadly, this experiment could give music teachers and students better insight into how music is learned and how to better report progress of learning.

What makes this system so interesting is not that it improves learning but that it does so with a passive measurement of cognitive load. Essentially, it is a tutor that uses introspection to adjust the difficulty of a given stimulus. Piano teachers cannot see how hard their students are working or how difficult something might be without having the student give self-reported information. Our system can measure students' cognitive load which would be a valuable tool for piano teachers and teachers in other domains.

A primary goal of music education is to shrink the gap between what what students are capable of learning and what teachers want students to work on. Many interfaces, multimedia platforms, and teaching methods have attempted to accomplish this by providing tools for students to learn and be evaluated as they practice [18]. These methods, however are still a reflection of self-reporting from students and judgements made by teachers. Our system not only can improve the learning process and lead to better performance but also does so without interruption from a teacher or the need for the student to make a decision of when to increase difficulty.

It is clear that cognitive load is a relevant measure for learning processes and should be taken into consideration even if it cannot always be directly measured. If piano teachers were to pay more attention to cues or examples of cognitive overload [10] instead of simply making judgements of how well they played a piece it could be beneficial to the student cognitively as well as improve how well they learn a piece of music. In fact, I believe it would be beneficial if teachers in any domain paid more attention to the cognitive demands of what students are learning and relied less on performance as a measure of how well something is learned. Performing well on a test or playing a piece of music well may not be a reflection of the quality of learning. For example, in our experiment, among participants who showed smaller or no differences in performance in the two conditions, they reported having a better "understanding" of the music or they at least felt that it was easier to learn in the BCI condition. This is likely because their cognitive load was kept at a more manageable level on average.

In addition to the importance of cognitive workload in learning piano, this study shows an interesting distinction between beginner and intermediate pianists. In our pilot studies we found that intermediate/advanced pianists showed no differences between conditions. When interviewed, these individuals mentioned that they didn't like the BCI condition and during the experiment there was very little time between the changes of difficulty suggesting that their workload was lower in general throughout the experiment. One possible reason the BCI condition was not as effective for more advanced pianists is that with enough experience cognitive overload is less likely to occur and thus does not need to be measured and adjusted for. As an advanced pianist I have experienced this in that there are very few times that I feel overwhelmed or overload when reading music for the first time. Perhaps our system could also be helpful in getting beginners to this cognitively comfortable state with music earlier if they were to use this system for a more extended period of time.

6.2 Future Work

The results of this study raise many questions and open the door for significant meaningful future work in this field. Beyond the scope of this experiment there is room for research in cognitive load musical BCIs, learning and cognitive load, and score following and musical assessment.

In the field of musical BCI applications there is very little done that adapts to the user's cognitive load other than BRAAHMS [1] and this study. It would be interesting to look at other manipulations with this type of brain measurement. The musical output was manipulated in the BRAAHMS study and the music notation was manipulated in this study, so why not manipulate the musical interface itself? There have been examples of software based instruments that can be changed and customized based on the need of the user [28] but no work has been done on manipulating these interfaces in real time. By measuring a user's cognitive load we could theoretically add more inputs and controls onto the instrument interface itself as their cognitive load drops. In theory someone using a device like this would start by creating some simple melody and gradual build up more controls and musical parameters to change as they become more comfortable with the current interface. This could help performers add and remove parameters to their interface more seamlessly without needing to dedicate energy to making decisions of when things should change.

As far as learning is concerned, there is a great deal of research that acknowledges cognitive load as an integral part to learning domains. However, much of this research highlights the need for an evaluative system for measuring cognitive load. It has been shown that fNIRS can do just this very effectively [1, 8]. It would be very interesting to see how cognitive load changes in different domains. One interesting area for research would be in more standard recall and memory tasks. It would be interesting to see if using a similar system of presenting stimuli based on cognitive load in a progressive manner would have the same benefits of performance on other tasks unrelated to music. Perhaps this should be tested on other motor control pattern tasks. Music is definitely a unique domain because reading music requires some amount of translation from visual stimuli to motor control which is why research in other domains would be useful to understand the benefits of measuring cognitive load in various learning environments.

Another area for study is score following and musical assessment measurements. What I found most interesting when researching score following techniques and related research is that there is not a standardized way to measure piano performance in a quantitative evaluation context. There are many methods of teaching piano but no definitive way to test the effectiveness of each with a standard measure. The reason most score-following software is not useful for this context is because the most robust of these programs are tempo dependent. They measure a user's accuracy relative to a predetermined speed of playing. This does not allow for someone to play at their own pace and gives no measurement of tempo consistency. Software that is not tempo dependent usually will not move on in the score until the correct notes are played, which forces the user to make corrections until they play correctly.

It is clear that this is an area for improvement in the domain of score following. There is no program that measures a user's accuracy, average speed, number of correct notes, number of incorrect notes, and number of extra notes. Not only would a program like this be useful in the context of research similar to this, but it could also be a helpful tool for music teachers who want to assess their students' progress in a more empirical manner. Rather than just making judgements based on watching and listening, they could augment their instruction with a systematic and consistent tool that provided objective measures of performance while allowing the student to play at their own pace with no interference in the way they play or read the music.

There are many implications of this research and a plethora of disciplines and designs to explore. This project was an important first step in better understanding these domains and provides valuable data and direction to any related work done in this field.

Chapter 7

Conclusion

In conclusion, we have successfully created a musical BCI that uses real-time fNIRS input as a measurement of cognitive load. We applied this BCI to a learning task for beginner pianists learning short pieces of music. We then ran an experiment evaluating the effectiveness of this system and showed significant improvements in learning with the BCI compared to a normal control. We affirmed our original hypothesis that using a BCI that alters the difficulty of music in real-time based on cognitive load improves performance and is a better learning experience for participants.

This study is a contribution to the HCI field in that it is one of the earliest to show the benefits of monitoring and utilizing cognitive load in a learning environment as well as being an innovation in music education practices and how musicians interact with the music that they are learning. Our system not only helped improve performance, enjoyment and ease of learning a piece of music but also provides an "introspective tool" to music educators who are otherwise unable to justifiably determine the current workload of their students. While it is unlikely that our system would be used in anything other than an experimental context, it still provides great insight into the effects of cognitive load on learning music and challenges traditional teaching methods.

The work done on the musical assessment portion of this project was also successful in identifying and implementing methods of objectively evaluating piano performance. The reported measurements and methods were substantive and provide a foundation for future work. Without these measurements, evaluation of our BCI would have been much less rooted in quantitative data.

This study is an early step towards improved music education, improved learning abilities based on cognitive load measurement, and better tools and methods of musical assessment.

Appendix A

Python Code

A.1 MIDI_Input.py

```
import pygame
import pygame.midi
from pygame.locals import *
from collections import Counter

pygame.init()

pygame.fastevent.init()
event_get = pygame.fastevent.get
event_post = pygame.fastevent.post

pygame.display.set_caption("midi test")
screen = pygame.display.set_mode((400, 300), RESIZABLE, 32)

pygame.midi.init()
count = pygame.midi.get_count()
for i in range(0, count):
        info = pygame.midi.get_device_info(i)
input_id = pygame.midi.get_default_input_id()
midi_in = pygame.midi.Input(1, 0)

print ("starting")

going = True
midi_list = []
time_list = []
tempo_list = []
```

```
note_off_list = []
note_list = []

while going:
        events = event_get()
        for e in events:
                if e.type in [QUIT]:
                        going = False
                if e.type in [KEYDOWN]:
                        going = False

        if midi_in.poll():
                midi_events = midi_in.read(1)
                note = midi_events[0][0][1]
                time = midi_events[0][1]

                if midi_events[0][0][2] != 0 and midi_events[0][0][0] != 128:
                        note_list.append(note)
                        if pygame.display.get_caption() != "midi working" :
                                pygame.display.set_caption("midi working")
                        if time_list and (time - time_list[-1]) < 150:
                                midi_list[-1].append(note)
                        else:
                                midi_list.append([note])
                                time_list.append(time)
                else:
                        note_off_list.append(time)

                midi_evs = pygame.midi.midis2events(midi_events,
    midi_in.device_id)

                for m_e in midi_evs:
                        print (m_e)
                        event_post( m_e )

print ("midi event list: ", midi_list)
print ("midi time list: ", time_list)
print ("midi tempo list: ", tempo_list)
print ("midi note list: ", note_list)
total_time = (note_off_list[-1] - time_list[0])
note_gap = total_time/len(time_list)
tempo = round(60/(note_gap/1000), 3)
note_confidence = []
notegap_total = 0
```

```
for i in range(len(time_list)-1):
        note_confidence.append(note_gap - (time_list[i+1] - time_list[i]))
        notegap_total += (note_gap - (time_list[i+1] - time_list[i]))

notegap_min = min(note_confidence)
notegap_max = max(note_confidence)
notegap_total = notegap_total/len(time_list)

print("number of beats played: ", len(time_list))
print("total time: ", total_time)
print("note gap:", note_gap)
print("average tempo: ", tempo, " BPM")
print("note gap range: ", notegap_min, " to ", notegap_max, " ms")
print("average note variance: ", notegap_total, " ms")
print("number of each note played: ", Counter(note_list))
print("total notes played: ", len(note_list))

print ("exit button clicked.")
```

A.2 variance.py

```
f = open('variance.txt')
f_new = open('variance.csv', 'w')

for line in f.readlines():

   participant = ''
   data = []
   variance = []
   i = 0
   number = ''
   data_text = ''

   while line[i] != ' ':
      participant += line[i]
      i+=1

   while i < len(line):
      if line[i] != ' ' and line[i] != '=' and line[i] != '[':
         number += line[i]
         if line[i] == ',' or line[i] == ']':
            data.append(int(number[0:(len(number)-1)]))
            number = ''
      i+=1

   for i in range(len(data)-1):
      variance.append((data[i+1] - data[i]))

   for i in range(len(variance)):
      data_text += str(variance[i]) + ','

   f_new.write(participant+','+data_text[0:(len(data_text)-1)]+'\n')
```

Appendix B

Music Stimuli

B.1 Training Task

Figure B.1: Example of a "difficult" piece used in training task

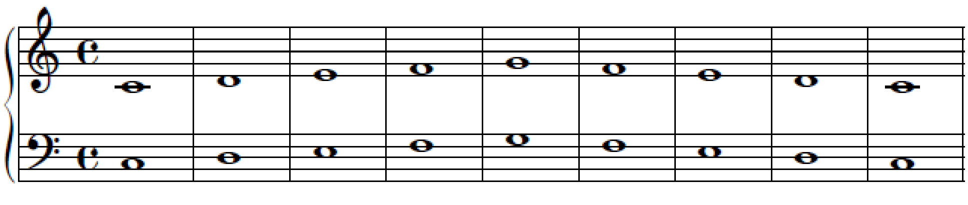


Figure B.2: Example of an "easy" piece used in training task

B.2 Learning Task

FIGURE B.3: Level 1 of score used in BCI condition in learning task

FIGURE B.4: Level 2 of score used in BCI condition in learning task

FIGURE B.5: Level 3 of score used in BCI condition in learning task

FIGURE B.6: Level 4 of score used in BCI condition in learning task

FIGURE B.7: Score used in control condition in learning task

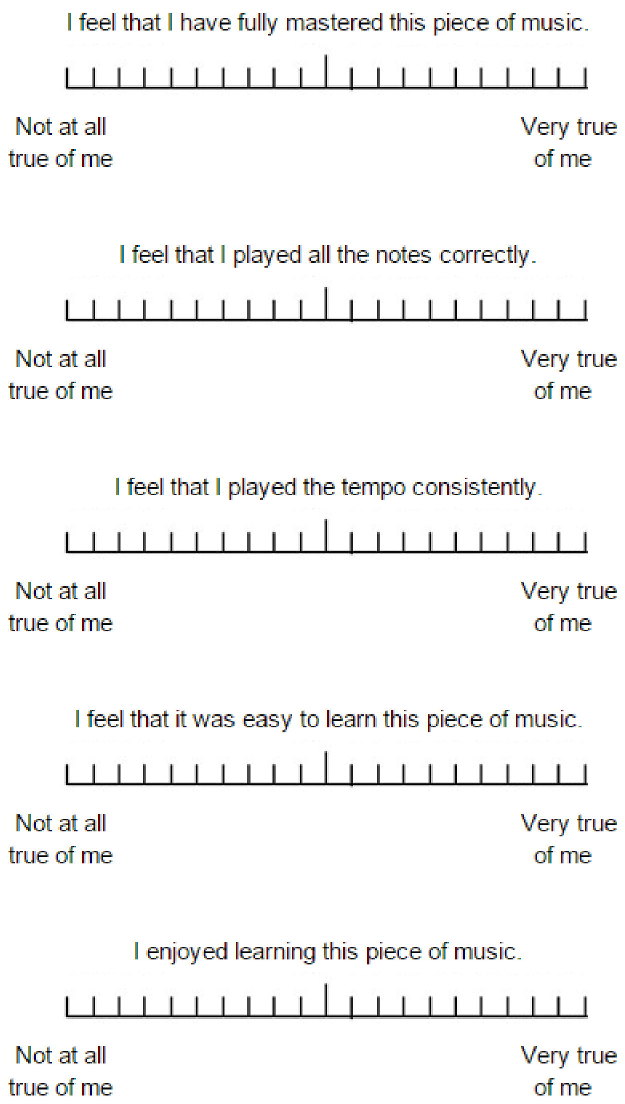
Appendix C

Questionnaires

C.1 Experiment Questionnaire

(to be inserted)

C.2 Post Experiment Questionnaire



I feel that I have fully mastered this piece of music.

Not at all true of me — Very true of me

I feel that I played all the notes correctly.

Not at all true of me — Very true of me

I feel that I played the tempo consistently.

Not at all true of me — Very true of me

I feel that it was easy to learn this piece of music.

Not at all true of me — Very true of me

I enjoyed learning this piece of music.

Not at all true of me — Very true of me

FIGURE C.1: Questionnaire given to participants after the learning task. Scale ranges from 0 to 20.

Appendix D

Musical Assessment Example

D.1 Python readout

```
midi time list:  [7061, 8547, 9833, 11079, 13143,
13673, 16024, 16669, 18733, 20051, 22050, 23924, 26270,
26509, 28367, 30406, 31074, 32567, 34093, 35527, 37303,
37819]
midi note list:  [55, 67, 74, 59, 59, 67, 79, 52, 59,
67, 74, 55, 60, 67, 48, 76, 50, 67, 74, 59, 74, 57, 67,
50, 66, 72, 60, 52, 50, 66, 60, 72, 72, 67, 55, 60, 71,
67, 55, 59, 74, 67, 59, 55, 74, 62, 59, 50, 52, 62, 69,
52, 60, 72, 64, 60, 52, 71, 67, 62, 55, 71, 61, 69, 64,
57, 66, 60, 54, 69, 55, 67, 71, 59, 60, 67, 59, 50, 62]
number of beats played:  22
total time:  33279
note gap: 1512.6818181818182
average tempo:  39.665  BPM
note gap range:  -838.3181818181818  to  1273.
6818181818182  ms
average note variance:  45.83264462809923  ms
number of each note played:  Counter({67: 12, 59: 9,
60: 8, 55: 7, 74: 6, 50: 5, 52: 5, 71: 4, 72: 4, 62: 4,
66: 3, 69: 3, 64: 2, 57: 2, 76: 1, 79: 1, 48: 1, 54: 1,
61: 1})
total notes played:  79
```

Figure D.1: Example of python code readout of a performance

D.2 Manual assessment

SUBJECT NUMBER

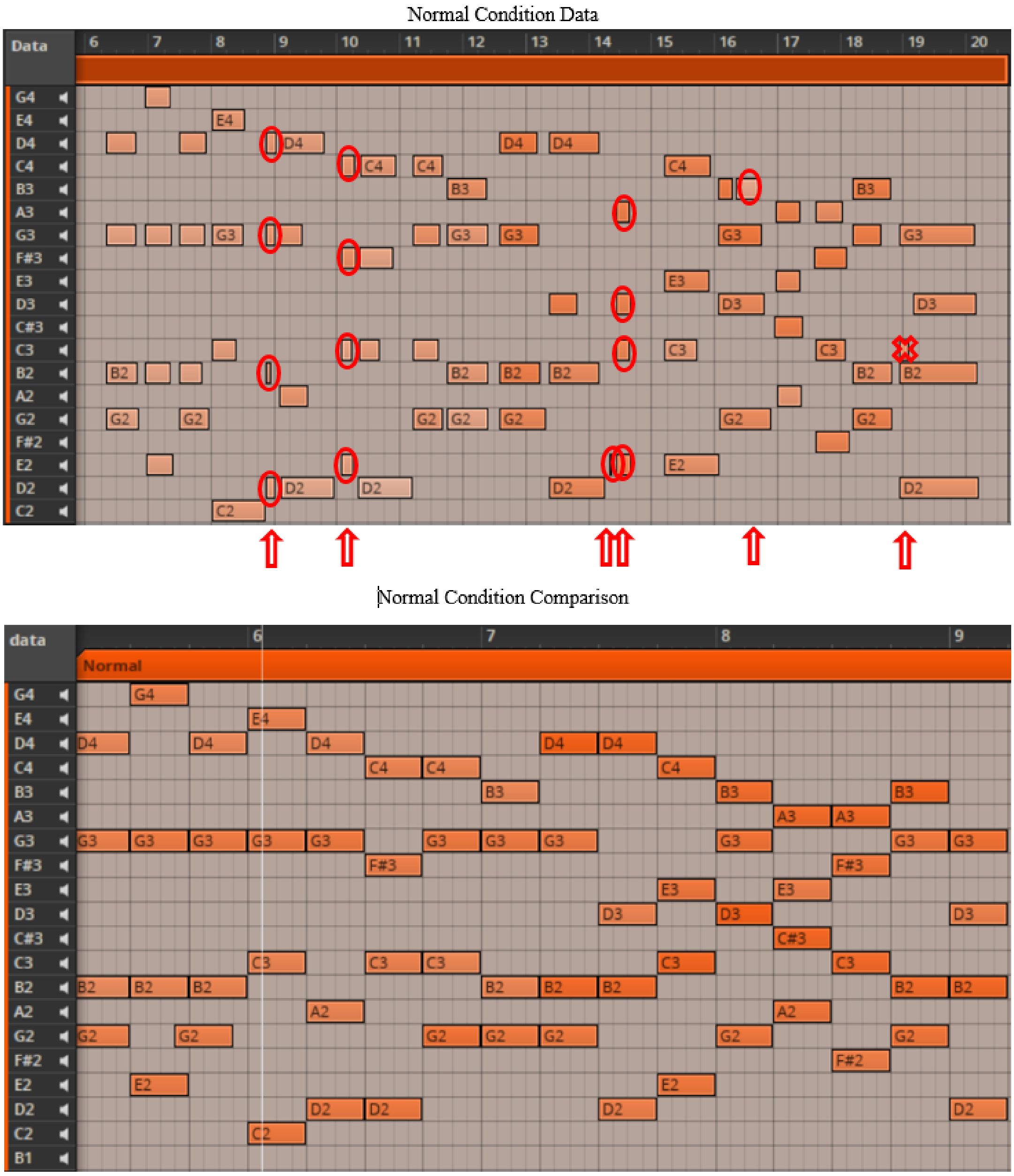


FIGURE D.2: Annotated manual assessment of musical performance. Extra notes circles, incorrect notes marked with 'X', errors indicated with arrows

Bibliography

[1] Beste F. Yuksel, Dan Afergan, Evan M. Peck, Garth Griffin, Nick Harrison, Lane Chen, , Remco Chang, and Rob J.K. Jacob. Braahms: A novel adaptive musical interface based on users' cognitive state. *Proc. NIME 2015, In Press.*

[2] Laurent George and Anatole Lecuyer. An overview of research on'passive'brain-computer interfaces for implicit human-computer interaction, 2010.

[3] J. Sweller. *Instructional Design in Technical Areas.* ACER Press, 1999.

[4] F. Paas and J. J. G. van Merrienboer. Variability of worked examples and transfer of geometrical problem-solving skills: A cognitive load approach. *Journal of Educational Psychology*, 86:122–133, 1994a.

[5] J. J. G. van Merrienboer and J. Sweller. Cognitive load theory and complex learning: Recent developments and future directions. *Educational Psychology Review*, 17(2): 147–177, 2005.

[6] Erin Treacy Solovey, Paul Schermerhorn, Matthias Scheutz, Angelo Sassaroli, Sergio Fantini, and Robert J K Jacob. Brainput : Enhancing Interactive Systems with Streaming fNIRS Brain Input. *Proc. CHI 2012*, 2012.

[7] Evan M. Peck, Beste F. Yuksel, Alvitta Ottely, Rob J.K. Jacob, and Remco Chang. Using fNIRS Brain Sensing to Evaluate Information Visualization Interfaces. *Proc. CHI 2013*, pages 473–482, 2013.

[8] Daniel Afergan, Evan M. Peck, Erin T. Solovey, Andrew Jenkins, Samuel W. Hincks, Eli T. Brown, Remco Chang, and Robert J.K. Jacob. Dynamic difficulty using brain metrics of workload. *Proc. CHI 2014*, pages 3797–3806, 2014.

[9] Grega Repovs and Alan Baddeley. The multi-component model of working memory: Explorations in experimental cognitive psychology. *Neuroscience*, 139(1):5–21, 2006.

[10] Richard E. Mayer and Roxana Moreno. Nine ways to reduce cognitive load in multimedia learning. *Educational psychologist*, 38(1):43–52, 2003.

[11] Ron J.C.M. Salden, Fred Paas, and Jeroen J.G. van Merrinboer. A comparison of approaches to learning task selection in the training of complex cognitive skills. *Computers in Human Behavior*, 22(3):321–333, 2006.

[12] Ron J.C.M. Salden, Fred Paas, Nick J. Broers, and Jeroen J.G. van Merrinboer. Mental effort and performance as determinants for the dynamic selection of learning tasks in air traffic control training. *Instructional Science*, 32(1-2):153–172, 2004.

[13] Fred Paas, Juhani E. Tuovinen, Huib Tabbers, and Pascal W.M. Van Gerven. Cognitive load measurement as a means to advance cognitive load theory. *Educational psychologist*, 38(1):63–71, 2003.

[14] Eduardo Miranda and Andrew Brouse. Toward direct brain-computer musical interfaces. *Proc. NIME*, pages 216–219, 2005.

[15] Eduardo Reck Miranda. Brain-computer music interface for composition and performance. *Int J Disabil Hum Dev*, 5(2):119–126, 2006.

[16] Yee Chieh Denise Chew and Eric Caspary. Museegk: a brain computer musical interface. In *Extended Abstracts CHI 2011*, pages 1417–1422, 2011.

[17] Olga Sourina, Yisi Liu, and Minh Khoa Nguyen. Real-time eeg-based emotion recognition for music therapy. *Journal on Multimodal User Interfaces*, 5(1-2):27–35, 2011.

[18] Graham Percival, Ye Wang, and George Tzanetakis. Effective use of multimedia for computer-assisted musical instrument tutoring. *Proceedings of the international workshop on Educational multimedia and multimedia education - Emme '07*, pages 67–75, 2007.

[19] J Chow, Haoyang Feng, Robert Amor, and BC Wnsche. Music education using augmented reality with a head mounted display. *Proceedings of the Fourteenth Australasian User Interface Conference*, pages 73–79, 2013.

[20] Sam Ferguson. Learning musical instrument skills through interactive sonification. *Proceedings of the 2006 conference on New Interfaces for Musical Expression*, pages 384–389, 2006.

[21] Gottfried Schlaug, Andrea Norton, Katie Overy, and Ellen Winner. Effects of music training on the child's brain and cognitive development. *Annals of the New York Academy of Sciences*, 1060:219–230, 2005.

[22] Steven M. Demorest and Steven J. Morrison. Does music make you smarter? *Music Educators Journal*, 87(2):33–39, 2000.

[23] Roberta W. Brown. The relation between two methods of learning piano music. *Journal of Experimental Psychology*, pages 435–441, 1933.

[24] S.F. Zdzinski. Measurement of solo instrumental music performance: A review of literature. *Bulletin of the Council for Research in Music Education*, 109:47–58, 1991.

[25] Sam Thompson and Aaron Williamon. Musical performance assessment as a research tool. *Music Perception: An Interdisciplinary Journal*, 21(1):21–41, 2003.

[26] Anna Jordanous. Score following: An artificially intelligent musical accompanist. Artificial Intelligence, School of Informatics, University of Edinburgh, 2007.

[27] D. Francis. Composing student learning. *University Teaching: academics' stories*, pages 131–137, 1997.

[28] Sergi Jord, Gnter Geiger, Marcos Alonso, and Martin Kaltenbrunner. The reactable: Exploring the synergy between live music performance and tabletop tangible interfaces. *TEI '07 Proceedings of the 1st International Conference on Tangible and Embedded Interaction*, pages 139–146, 2007.