"This one goes to eleven…": A Methodological Study of the Recording and Evaluation of Emotional Response to Music

Nicholas Shortway

Submitted January 2007

Music Technology Area

Department of Theory

Schulich School of Music

McGill University

Montreal, Quebec, Canada

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Master of Arts

Your file Votre référence
ISBN: 978-0-494-32563-6
Our file Notre référence
ISBN: 978-0-494-32563-6

Abstract

The distinction between emotions *expressed* by and *evoked* by music is often blurred by the use of methodologies that are difficult to compare, preventing progress towards a unifying theory of emotion and music. In this thesis, a methodology for experimentation that clarifies how research in these areas can be conducted clearly and independently is proposed. The study of evoked emotions is emphasized, and two novel controllers are developed to examine methods of recording emotional response continuously. These controllers, along with a number of the proposed methodological changes, are tested against an established controller in an experiment designed to record emotional changes to participant-selected musical pieces. The results support predictions regarding the effects that the experimental setting can have on the emotional responses of the participant. Usability ratings of one of the new controllers were found to be slightly higher than those of the established controller, while providing an interface that is less emotionally distracting for the participant. It also provides the ability to record instances of physiological reactions evoked in the participant.

Sommaire

La distinction entre les émotions *exprimées* par et *évoquées* par la musique est souvent rendue floue par l'utilisation de méthodes difficilement comparable, empêchant l'avancement vers une théorie unifiante de l'émotion et de la musique. Dans cette thèse, une méthode pour l'expérimentation qui clarifie la manière dont la recherche dans ces domaines peut se conduire clairement et indépendamment est proposée. L'emphase est mise sur l'étude des émotions évoquées et deux nouveaux contrôleurs sont développés afin d'étudier les méthodes d'enregistrement continu de la réponse émotionnelle. Ces contrôleurs, ainsi qu'un certain nombre de propositions de changements méthodologiques, sont testés contre un contrôleur existant dans une expérience conçue pour enregistrer les changements émotionnels provoqués par des pièces musicales choisies par les participants. Les résultats étayent les prédictions concernant les effets que la situation expérimentale peut avoir sur la réponse émotionnelle des participants. Des évaluations de l'utilité d'un des nouveaux contrôleurs étaient légèrement supérieures à celles du contrôleur existant, tout en fournissant une interface qui distraie moins le participant en termes de sa réponse émotionnelle. Il présente également la possibilité d'enregistrer des instances de réactions physiologiques évoquées chez le participant.

Acknowledgements

First, I want to thank Stephen McAdams for his tireless assistance towards the completion of this thesis. His knowledge, enthusiasm, and patience throughout the entire process were the catalysts that kept this project on track. I would also like to thank Mark Ballora for inspiring me to forgo a lifetime of mind-numbingly boring IT work to pursue a field that I didn't even know existed. Last, but certainly not least, I want to thank my family: Mom and Dad for their endless support for every one of my admittedly impromptu endeavors, Chris for always being there to give me hope and put things into perspective, and Windee for her ingenious help out of many statistical analysis meltdowns. Without all of you, I would not be here today, so thank you all once again.

Table of Contents

List of Tables vi
List of Figures vii
1. Introduction 1
  1.1 Overview 1
  1.2 Relevance of Project within the Field 3
2. Cognitivism vs. Emotivism 5
  2.1 Defining the Terms 5
  2.2 Differing Research Goals 6
  2.3 Ineffability and the Emotional Experience of Music 7
3. Experimental Design Considerations 13
  3.1 Goals 13
  3.2 Participant Selection 13
  3.3 Stimulus Selection 15
    3.3.1 Style/genre 15
    3.3.2 Familiarity 17
    3.3.3 Length 18
  3.4 Experimental Environment 20
  3.5 Instructions 24
  3.6 Questionnaires 28
    3.6.1 Pre-experiment questionnaire 29
    3.6.2 Mood assessment 31
    3.6.3 Post-experiment questionnaire 32
4. Interface Design Considerations 36
  4.1 Preliminary Research 36
  4.2 Pilot Experiment Methods 36
    4.2.1 Participants 36
    4.2.2 Stimuli 36
    4.2.3 Equipment 36
    4.2.4 Procedure 37
  4.3 Pilot Experiment Results and Discussion 37
  4.4 Early Continuous Response Interfaces 39
  4.5 One-Dimensional Digital Continuous Response Interfaces 41
  4.6 Two-Dimensional Digital Continuous Response Interfaces 43
  4.7 New Interface Designs 45
    4.7.1 Transparent Emotional Encoding 45
    4.7.2 Interface Size 46
    4.7.3 Rating Scale Attributes 46
5. Main Experiment Design 49
  5.2 Experiment Methodology 52
    5.2.1 Participants 52
    5.2.2 Stimuli 52
    5.2.3 Equipment 53
    5.2.4 Procedure 53
6. Results 54
  6.1 Results 54
    6.1.1 Music Education 54
    6.1.2 Music Listening Habits 54
    6.1.3 Mood Assessment 55
    6.1.4 Experimental Procedure Data 56
    6.1.5 Joystick 56
    6.1.6 Button 58
    6.1.7 Slider 59
    6.1.8 Controller Comparison 61
    6.1.9 Final Questionnaire 62

7. Discussion 64
7.1 Introduction 64
7.2 Questionnaires 64
7.3 Experimental Environment 67
7.4 Controllers 68
7.4.1 Joystick 68
7.4.2 Button 70
7.4.3 Slider 71
7.4.4 Cross-Controller Comparison 73
8. Conclusion 76
Appendix A: Questionnaires and Instructions 78
A.1 Pilot Experiment Questionnaire 78
A.2 Background Questionnaire 79
A.3 Joystick Instructions 80
A.4 Button Instructions 81
A.5 Slider Instructions 82
A.6 Post-Controller Questionnaire 83
A.7 Post-Experiment Questionnaire 84
A.8 Research Ethics Approval Form 85
Bibliography 86

List of Tables

Table 1: Experimental Design Considerations 35
Table 2: POMS Results 56
Table 3: Frequencies of Controller Presentation Order 56
Table 4: Joystick Statistics 57
Table 5: Joystick Correlations 58
Table 6: Button Statistics 58
Table 7: Button Correlations 59
Table 8: Slider Statistics 60
Table 9: Slider Correlations 60
Table 10: Controller Data Interaction Correlations 61
Table 11: Chill Response Correlations 62
Table 12: Experimental Setting Responses 63
Table 13: Comfort Rating / POMS Correlations 63

List of Figures

Fig. 1: Circumplex model of affect, reproduced from Russell (1980) 8
Fig. 2: Responses to *La Boheme,* reproduced from Madsen (1997) 27
Fig. 3: Responses to *Symphony no. 104,* reproduced from Madsen (1998) 27
Fig. 4: The Continuous Response Digital Interface 41
Fig. 5: CRDI musical tension overlay, reproduced from Madsen & Frederickson (1993) 42
Fig. 6: Joystick Controller 49
Fig. 7: Button Controller 51

1. Introduction

1.1 Overview

The purpose of this project is to examine the current methodology used with experiments involving the continuous tracking and recording of emotional responses to music and to develop a set of rigorous guidelines that researchers must follow in order to collect the most useful and valid data possible. In addition to analyzing and improving upon the current methodologies used for tracking and recording emotional response, two physical interfaces are developed that attempt to avoid the pitfalls commonly found in currently used interfaces, and their usability is compared to a previously established interface. With these goals in mind, exploration of the fields of emotional response and continuous response is essential to determine the most productive changes to the current methods.

The first step is to examine the major theories about emotion and how they have been studied within the realm of music research. This background gives the project a basis upon which to examine the varied implementations of continuous response systems used for tracking emotional response. All the relevant systems must, in some way, have roots established within the basic theories of human emotional response to be valid for experimental use.

Following from the research into emotion in music, a clear need to differentiate between the two varieties of "emotional response to music" becomes evident. These two variants are confused and confounded throughout most of the experimental literature that is available on the topic of continuous measurement of emotional response to music. The supposed "poles" of the musical/emotional experience, originally defined by Kivy (1990) as "cognitivist" and "emotivist", have been mixed up within most aspects of the experimental designs on a fairly regular basis. Due to the extreme importance of their discrimination within experimental design, this project defines the terms, their similarities, their differences, and their implications within these particular types of experiments in Chapter 2. This chapter also discusses the ineffability of the emotional experience of music and introduces the terminology and techniques used in the thesis' experiments in order to embrace this experiential quality.

In order to develop a set of guidelines for designing experiments involving emotional response to music, each individual aspect of experimental design is broken down, implementation in past experiments is examined, and the necessary changes to such elements are proposed, in an effort to bring uniformity to experimental organization. This uniformity allows the data to be compared and analyzed across multiple studies. The effect of music on human emotions is almost infinitely complex, so there must be a standard framework within which these studies are conducted. Once this framework is solidified and validated, progress towards a unified theory of emotional response to music can proceed with minimal obstacles. This proposed framework is outlined in Chapter 3.

In addition to the proposed methodological changes, the design of the physical interface used by the participants to make their continuous emotional ratings is a key factor in a successful experiment. In Chapter 4, a survey of the interfaces used in these studies is offered, discussing both their mechanical design and their interconnection to the instructions given to the participant. The interaction between the participants and the interface is examined, pointing out problems that develop with certain designs. The chapter finishes with a discussion of the design implications considered during the development of the novel interface implemented within this project.

To begin validation of the proposed experimental framework, an experiment involving a subset of the suggested modifications is conducted. While an attempt to validate every aspect is in order, the scope of this thesis restricts the amount of experimentation and thus will require an appropriately selected subset of changes that can be validated. This experiment also tests the new physical interface against interfaces used in prior emotional response studies. The defining attributes of the interface are clearly unique from those of the prior systems, so an overt preference of the user for a particular system presents itself. The specifics of setting up and running this experiment are found in Chapter 5.

With all the data collected, the results are presented in Chapter 6 and discussed in Chapter 7, pointing out how the proposed framework and its corresponding physical interface compare to previous systems when the goal is to record accurately the movement of a participant's emotional experience throughout a piece of music.

Finally, the work done in the thesis is summarized and future directions for the field of evoked emotional responses to music are proposed in Chapter 8.

1.2 Relevance of Project within the Field

Why do humans listen to music? It is certainly not to say inwardly, "this composer was trying to convey sadness" or "it sounds happy, so I probably should feel happy." We humans listen in order to experience emotions, to get chills down our spines, to recall an emotional memory, to feel connected to art and its meaning. It is only natural to strive towards an understanding of that connection, even if it differs from person to person. Yet, within the realm of music research, this connection between evoked emotion and music has only been recently brought into the spotlight.

For decades, scientific researchers have avoided the question of evoked emotions in response to music. There have been countless numbers of articles, theses, and lectures on the topic of perceived musical emotion, all generating similar data showing that participants (within a culture) agree almost 100% of the time regarding the emotions that one perceives within a piece of music, i.e. emotions that one perceives the music as trying to convey. Researchers have used words (Hevner, 1936), rotary dials (Madsen, Brittin, & Sheldon, 1993), and two-dimensional computer interfaces (Tyler, 1996; Schubert, 1999) to determine what emotions are expressed by music, but very few have dared to ask the question, "What are the emotions that people actually experience when listening to music?" It's not a simple question and I am not ready to answer it. It has been avoided for years due to the vast number of variables that complicate the answer to it. Only in the last few years have researchers begun to breach the topic in a more systematic fashion (Sloboda, 1991; Gabrielsson, 2002; Nagel, Grewe, Kopiez, & Altenmüller, 2005). Many will argue that the topic has not been examined thoroughly because there will be no concrete answers; there will never be hard and fast rules about experienced emotional responses to music. This is true from a universal standpoint. One song will never *make* everyone sad in the way that one song will unanimously be *perceived* as sad. Does this make the question less valid? Even though we may not be able to develop a small set of all-inclusive rules governing what types of music evoke which emotions in every listener, is it not important to establish a standard way of collecting and analyzing these different emotional responses, regardless of how personal and individualistic they may be?

Once research into evoked emotion is undertaken, and once the sheer complexity of the phenomenon is recognized, it becomes evident that a mere handful of studies will not reach a cogent resolution. This is where a systematic, well-organized methodology bridges the gaps between future studies to keep the data comparable across projects. Without a rigorous standard applied to the experimental configurations and procedures, the data is doomed to be difficult, if not impossible, to compare with previous studies, therefore lacking momentum towards any kind of unified theory of emotional response to music. For example, when one study asks the participants to "move the dial corresponding to your aesthetic response" (Madsen et al., 1993, p. 61) and another asks the participants to "move the mouse cursor on the monitor screen to correspond to your responses of relaxing/exciting and ugly/beautiful" (Madsen, 1998, p. 549), the responses cannot be reliably compared even though the studies were meant to examine the same phenomenon. The ambiguities in the wording of the instructions prevent any resulting comparison from being valid. The development, and universal acceptance, of a novel methodology is the rigid backbone that will allow for the much-needed growth in this field of research.

Research into evoked emotion is the door into the basis for the adoration of music and its arcane effects on the human psyche. However, the only way to open this door is to have wide support for a rigorous methodology that can unify the research by making it fully compatible for analysis.

2. Cognitivism vs. Emotivism

2.1 Defining the Terms

In the last 16 years, since Kivy adopted the terms from their philosophical roots in his book entitled *Music Alone* (1990), the use of the characterizing words "cognitivist" and "emotivist" has been mutated, distorted, and at the very least inappropriate. The original definitions of the terms are as follows:

> Those I am calling musical emotivists believe that when, under normal circumstances, musical critics, theorists, or just plain listeners call a piece of music "sad," it is because it makes us sad when we listen to it; and what they mean by "sad" music, I will assume, is music that normally arouses sadness in the normal listener. The musical cognitivists, like the emotivists, believe that it is proper sometimes to describe music in emotive terms. But unlike the emotivists, they do not think that sad music is sad in virtue of arousing that emotion in listeners. Rather, they think the sadness is an expressive property of the music which the listener recognizes in it, much as I might recognize sadness as a quality of a dog's countenance or even of an abstract configuration of lines. (p. 146)

The problem that arose during the interpretation of these definitions was the belief that the terms define the possibilities of a musical experience, yet Kivy was not saying that people do not experience emotions when listening to music. His entire reason to define such terms was to clear up the discrepancies between the differing methods of categorizing music and its expressiveness. He follows the cognitivist perspective, believing that music expresses emotions such as joy, sadness, anger, etc., but that the music does not make us feel those same emotions, due to a lack of an object towards which the emotions may be directed. His belief does not end there, as many believe it to, because he went on to say that while music does not induce the expressed emotions, it can "move" us emotionally, as we've all experienced at one time or another.

As mentioned earlier, these two words have been distorted in their usage within this field of research. For example, as recently as 2001, Scherer and Zentner wrote that, "Whereas 'emotivists' hold that music elicits real emotional responses in listeners, 'cognitivists' argue that music simply expresses or represents emotions" (p. 361). This generalization is a common one that widens the breach between the two perspectives, when the true perspectives are actually closer than most would think. This can be argued by the simple fact that every single person has experienced emotions in response to music at one time or another. Call it 'being moved', 'feeling the music', 'witnessing the beauty', or any number of similar phrases, but it all comes down to the fact that music evokes a feeling in humans that is often ineffable. Kivy was extremely meticulous in his definition of emotivists: they do not argue that music simply evokes emotions. They argue that music *evokes* the particular emotion that is *expressed* in the music, and that is why that certain excerpt is referred to by the name of that emotion, e.g. if it evokes sadness in a person, it is 'sad' music. This definition, in its strictest sense, does not leave much room for interpretation, which led many researchers to jump to conclusions about the uses of the terms. What would one call a person who finds (through personal experience) that a certain piece of music that *expresses* sadness actually *evokes* an ecstatic feeling? They would not be an emotivist because of the negative correlation between the expression and the evocation. They would actually lean towards the cognitivist perspective because they identify the expression of the music as one of the 'everyday emotions', and have come to the conclusion that the *expressed* emotion is not the same as the *evoked* emotion. This fundamental problem with the terminology forced the mutation of the terms within the literature. Yet, if a researcher uses the mutated forms of the terms, they are left with cognitivists believing that emotions will never be evoked by music, which was never the claim in the first place: "I do not claim that the second movement of *Eroica* does not *move* us emotionally, only that it does not move us to *sadness*, its predominant emotion" (Kivy, 1990, p. 147). So, if both sides of the 'argument' agree that there is some kind of evoked emotional response to music, researchers must abandon these terms, unless using them to simply describe the goal of their research: 1) to analyze evoked emotion in the participants (emotivist study) or 2) to analyze the expressed emotion in the music (cognitivist study). If these terms are used solely to define the type of research being done, and the rigorous standards proposed in this thesis are followed, they have the potential to keep the goals of each successive study as evident as possible.

2.2 Differing Research Goals

First, it must be made clear that it is not the purpose of this project to discredit any research being done with the cognitivist goal of analyzing expression of emotional content in music. There are many benefits of this research, including the understanding of what musical features are responsible for the perception of certain emotions expressed in the music. The importance for the research methodology is that a cognitivist study cannot take place in the same experiment as an emotivist study. If a researcher begins a study without making this distinction concerning the goal of the project, the results will most likely be completely useless, because, as will be discussed as this thesis progresses, every single parameter involved in the experimental design has the potential to push the data towards either cognitivism or emotivism.

The explicit goal of the experiment contained within this thesis is emotivist; it will attempt to accurately track the emotional force, a term discussed later, experienced by participants as they listen to familiar pieces of music. The methodology that surrounds the experiment, however, is discussed in terms of both cognitivist and emotivist experiments in an attempt to keep the considerations for each experimental design procedure as straightforward as possible.

2.3 Ineffability and the Emotional Experience of Music

Try to recall an intense emotional experience you've had in response to music. Try to recreate it in the mind. Try to feel the surges of chills up and down your spine. Try to remember the 'lump in your throat' or the tearing in your eyes. Try to remember the sonic events that triggered the sensation: the melodies, harmonies, timbres, and rhythms. Now, try to speak aloud the emotion you felt. Was it sadness? Was it happiness? Was it wistfulness? Was it a combination of these? Most people will balk at the attempt to verbalize such a feeling, and usually revert to something along the lines of, "I can't really describe it." This is a completely normal reaction to such an experience, because intense emotional experiences will often defy any type of categorization or label placed on them. Sloboda remarks that, "Although these experiences have an emotional component, they are hugely more complex than the 'garden variety' emotions of happiness, sadness, etc. in ways that are not fully explored" (2002, p. 245). Music also draws on feelings that we don't often experience in quotidian life: "emotions induced by music and emotions felt in everyday situations differ in their relative frequency of occurrence" (Juslin & Zentner, 2002, p. 10). Taking this view into account, it seems curious that so many studies have tried to pigeonhole the emotional experience into systems that have only been designed for the categorization of everyday emotions. One such system was developed by Russell (1980). It rated a large set of emotions on a two-dimensional grid, now known as the "circumplex model of affect" (Fig. 1).

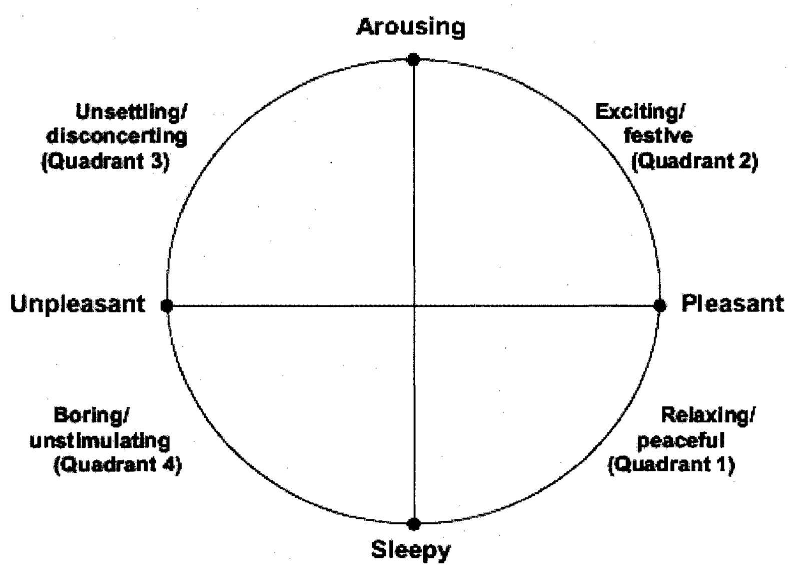

Fig. 1: Circumplex model of affect, reproduced from Russell (1980)

The dimensions of the model are valence (x-axis) and arousal (y-axis). Russell found that most emotions could fit within this two-dimensional space very simply, and there was very high agreement across participants regarding the location of each emotion in the space. This model has been used in a number of cognitivist studies (Tyler, 1996; Schubert, 1996; Schubert, 1999), within which it has been shown to report accurately the emotions expressed in the stimuli. However, musical expression of emotions does not cover all the emotions that people have felt when listening to music. People do not often perceive a piece of music as *expressing* emotions like nostalgia and awe (both being emotions found by Zentner, Meylan, and Scherer, 2000, as having a much higher frequency of occurrence in musical emotion evocation than everyday situations). Yet, when the circumplex model of affect is used in an emotivist study, as in Nagel et al. (2005), where does a participant rate an emotion like nostalgia within the two dimensions? More importantly, how does the participant rate an emotional experience that they cannot even analyze enough to allow the most modest verbalization? Scherer and Zentner (2001) suggest that the answer to this problem is in the development of a new taxonomy of musically aroused emotion:

> [This taxonomy development] requires a theoretical background that does not prejudge the issue, as is the case with emotion theories that either, like discrete emotion theories, start from the assumption of a limited number of basic or fundamental emotions or that, like dimensional theories, focus exclusively on the valence and arousal dimensions of emotional feeling. (p. 381)

They both agree that no matter how proven and substantiated a model of perceived emotion is, it will fall short when it comes time to describe accurately even a single feeling produced in a listener. However, while the idea of inventing a new taxonomy sounds like the best choice, one unbiased by previous models and as open as possible to the wide spectrum of emotional experiences related to music, it very possibly will never be able to envelop the inherent subjectivity of these responses. The only way to bypass such problems is to embrace the ineffability of the experience.

Verbal self-report brings with it myriad complications that can very easily compromise the integrity of a research study. First, as described above, the participant may simply not be able to put into words the phenomenon that was just experienced. One of two possibilities will be found in this case: either the study will force the participant to take the indescribable experience and strip it down to fit within the constraints imposed by the experiment, or the participant will be allowed to express the experience in their own words ad libitum. So, the researcher will be left with either too little/wrong information or so much information that any kind of inter-participant comparison is completely impossible. Second, depending on the environment of the experiment, social effects may influence the verbal response. For example, if a participant hears a piece of music that reminds them of an unpleasant time in their life, they may modify their self-report in order to keep their personal experience from being announced to other participants that may be present or even to the researchers themselves. Third, participants can easily feel, especially in the context of an experimental setting (i.e. lab, testing equipment, looming researchers), that they need to answer a question 'correctly'. Orne (1959) devised the term "demand characteristics" to represent all the cues in an experiment that reveal the experimental hypothesis. The need to answer 'correctly' is because of a desire to play "the role of a good experimental subject" (Orne, 1962) in the terms of the participant's perception of the demand characteristics. Orne also found that this need can exist on a nonconscious level. Therefore, demand characteristics may draw the participants away from their own emotional response and towards the easy-to-recognize *perceived* emotions they were able to ascertain from even a brief listening session. However, if the researcher abandons this seemingly futile quantification attempt, it is possible to analyze the power of the experience without forcing labels upon it.

McAdams et al. (2004) defined a term called "emotional force" that was a simple rating scale representing the force of the emotions experienced by the listeners, regardless of the type of emotion. This scale, by not imposing any categories or labels on the varied responses, allows for quantification of the emotional experience on the basis of its strength. The simplicity of the emotional force scale is its main merit. While other scales used in similar studies, such as "aesthetic response" and "felt emotional response" (discussed later), are easily left open to interpretation, "emotional force" solidifies the goal of the measurement scale for the listener with almost no explanation needed. With "aesthetic response", participants may be confused along the cognitivist/emotivist perspective, because the phrase may be interpreted as the aesthetic quality recognized in the music, the response that the so-called aesthetic quality elicits in themselves, or some combination of both sides. "Felt emotional response", on the other hand, is rarely taken under the cognitivist perspective, but it may confuse the participant because it is ambiguous about whether they should be encoding the general feelings evoked, or whether they should be trying to coerce these feelings into a preconceived subset of emotional states. The words "emotional force", especially when preceded by "your", have little chance of being misinterpreted. In addition to this feature, the emotional force scale does not suppress the heterogeneous nature of the emotional response to music. By not requiring that the participants try to use words to describe an experience that is naturally ineffable, they do not feel pressure to be 'correct' with their self-report answer. Not only does this ensure that the recorded responses are as genuine as possible, it also allows the participant to feel more comfortable within the experimental setting, thus letting the music's emotional impact be of the highest magnitude possible. Lastly, the scale is extremely well suited to being implemented within a continuous response framework. Unlike verbal self-report, which can only be used at a small number of intervals throughout the audition of a stimulus (which may already impede the flow of the emotional response), the emotional force scale can be rated by a participant seamlessly, following the ebb and flow of their emotions. Due to these benefits, the experiment in this thesis will employ the emotional force scale for all ratings on the controllers involved.

The other facet of the emotional experience that will be recorded within this experiment is the experience of physiological reactions in response to the music. Often referred to as "thrills" (Goldstein, 1980; Sloboda, 1991) or "chills" (Panksepp, 1995; Nagel et al., 2005; Grewe, Nagel, Kopiez, & Altenmüller, 2005), this relatively distinct experience produces one or more physiological reactions in the listener, and is therefore reasonably easy to track and record over time. Sloboda supports this claim, stating that, "[thrills] have the benefit of being stereotypical, memorable, clearly differentiated from one another, and easily identifiable" (1991, p. 110). Technically, the use of the term "thrills" is more accurate than "chills" because the latter is a descriptive word for one type of the former. "Thrills" is an encompassing term that describes any kind of experienced physiological change in response to music, e.g. chills down the spine, lump in the throat, tears, trembling, sweating, etc. In common usage, though, both Sloboda (1991) and Panksepp (1995) found that "chill" is used much more often, mainly because that particular physiological event occurs more often than any other in response to music. Not only is it the primary response to music, but Goldstein found that 97% of listeners within his study declared music as the cause of this type of feeling in any situation. The source of these "chills" from a biological standpoint is still a bit of a mystery, but has often been linked to the autonomic nervous system (ANS) (Blood & Zatorre, 2001; Panksepp & Bernatzky, 2002) and is theorized to be directly connected to human emotional processes (Sloboda, 1991). Recent studies (Lowis, 1998; Nagel et al., 2005; Grewe et al., 2005) have also found a correlation between recorded experiences of "chills" and the participants' heart rate and galvanic skin responses (GSR), a measure of conductivity of the skin that is related to the arousal level of the ANS. These direct physiological measures are more fully discussed later in the thesis. It is only essential at this point that the theory behind the experience of "chills" and its importance within the experience of emotional response to music be recognized.

With the difference between cognitivist and emotivist studies clarified, progress can be made within each research direction, which is emotivist in the context of this thesis. Therefore, having the evoked emotional response in mind, the researcher must employ experimental techniques that will not underestimate the ineffability of the actual experience. This consideration is the impetus behind the utilization of the "emotional force" rating scale devised by McAdams et al. (2004) within the experiments in this thesis. It is also clear that an important, and relatively simple-to-track feature of evoked emotional responses to music are "thrills," or varied physiological responses linked to the ANS. The high frequency of the presence of "thrills" in the human response to music necessitates their inclusion in emotivist experiments, thus they are recorded alongside emotional force in the experiments discussed later. Nevertheless, an experiment cannot take place without careful consideration of the impact of its design on the participants and their understanding of the demand characteristics. It is evident from the vast confusion present in the cognitivist/emotivist debate that a simple distinction can easily be taken out of context, even for researchers, so clarity and rigor in the experimental design are paramount.

3. Experimental Design Considerations

3.1 Goals

In this chapter, a set of comprehensive guidelines is offered to ease the design of an emotional response study. Following the guidelines is important for a study with an emotivist goal, but many of them apply to cognitivist studies as well, and result in further support for the dichotomy of the research goals. The guidelines are presented chronologically with the process of the experiment, beginning with deciding on the preferred characteristics of the participants and progressing through to the final questionnaires.

3.2 Participant Selection

Clearly, there are many factors to consider when choosing a set of participants for an emotional response study. Variants such as musical background, listening habits, and musical preference can all have serious effects on the outcome of the experiment. The depth of the effect that participant selection has on the experiment is directly related to the other factors in the experimental design as well, such as stimulus selection, experimental environment, and instructions.

One of the main criteria used in participant selection for emotional response experiments tends to be musical training. Researchers will select equal groups of musicians and nonmusicians, or music majors and nonmusic majors, (Madsen & Frederickson, 1993; Madsen, Byrnes, Capperella-Sheldon, & Brittin, 1993; Lychner, 1998) and believe that this distinguishing characteristic between the participants will produce meaningful distinctions in the resulting data. This approach may be useful in studies that have strictly controlled stimulus selection, e.g. knowing for a fact that the musicians have heard a certain piece of music many more times than nonmusicians and testing the difference in emotional desensitization, yet is not always an appropriate criterion on which to select participants. The main problem with this selection routine is the assumptions it makes about the participants' musical backgrounds. Knowing that a participant has played violin since they were 4 years old or that a participant has never stepped foot inside a music class is inconsequential evidence for a participant's relationship with music (or lack thereof). While it may be more likely for the former to have a deeper intimacy with music, one cannot make this assumption based on education or training alone. Also, selection based on musical training tends to be ineffectual when considering both the source of emotional evocation and the intensity of the experience. Bigand (2003) showed that nonmusicians' and musicians' ability to recognize and process subtle structural patterns in music, which are maintained as a principal cause of the evocation of emotion (Sloboda, 1991), was not determined by a participant's musical training if the task did not call upon explicit knowledge derived from formal training. In terms of intensity, Gabrielsson (2001) found that musicians and nonmusicans reported similar levels of intensity while recalling their strongest emotional reactions to music. In cognitivist studies, this is somewhat of a moot point, considering that perception of emotion in music is relatively uniform, even across cultures (Krumhansl, 1997). Yet, for emotivist studies, the kinds and levels of emotions induced in the participants is going to depend greatly on variables outside of the realm of musical training.

In a study involving musical preference, Rentfrow and Gosling (2003) found that a person's choice of music depends highly on a vast number of factors including their personality, self-esteem, level of rebelliousness, current mood, current level of physiological arousal, and cognitive abilities. Therefore, if a researcher simply selects and groups participants on their level of musical training, the data has the potential to be fragmented without having at least a partial understanding of the factors discovered by Rentfrow and Gosling. Simple questions posed to the potential participants can be used to fill in the extra data and ensure that the person fits with the desired profile. These preliminary questions can help to determine if a participant's information in the questionnaires will be relevant later for the scope of the experiment. Depending on the goal of the experiment, the types of questions may differ. For instance, if the aim of the study is to analyze the connection between the structure of music and elicited "thrills", one might ask potential candidates whether they regularly experience any kind of physiological change in response to music. The goal is to determine what structural features of music have the ability to cause these effects, so it would be unwise to have participants who do not experience "thrills" included in the participant population. The answer to that question will provide a better source of participants than picking a sample based on whether or not their prospective degree is rooted in music.

There are also times at which simple questions can be used to find the participants that will accurately represent both sides of a hypothesis. Consider a study that examines the effect of trance music on mood states. The researchers should inquire about each prospective participant's opinion about trance music, being the style in question. Then, if the participant pool is comprised of an equal number of participants who enjoy listening to trance music and who do not, results showing that a large percentage of the participants had similar changes in mood state can be deemed as universal regardless of individual music preference.

The common considerations, such as age, gender, handedness, etc. all apply, and should be handled in a similar way to most other studies unless the goal of the study is specific to one of the groups. Having an even distribution of these variables is preferable, while taking into account the caveats involved with musical preference, personality, and listening habits.

Participant selection practices outside of the standard selection of musicians and nonmusicians can be very beneficial for the outcome of an experiment. The most important aspect is to stay away from assumptions regarding musical background (both listening and education). This attention to musical background is especially important in an emotivist study because of the large number of factors affecting the evoked emotional response.

3.3 Stimulus Selection

Selection of the musical stimuli, like participant selection, is highly dependent on the overall goal of the research. Relevant selection stipulations are divided into three main categories: style/genre, familiarity, and length.

3.3.1 Style/genre

Musical style (or musical genre) is one of the features of experimental design that has changed the least over the recent years. Most studies have avoided using any music beyond instrumental classical music for a variety of reasons, all which have their merits, yet end up limiting the scope of application of the research. It is true that classical music keeps analysis simple, with almost every piece being reproducible directly from a score, musical structure that follows established guidelines, somewhat predictable peaks and valleys in tension, and with about half of the participant pool in each experiment listening to this genre on a regular basis (another reason some researchers may choose the musician versus nonmusician participant selection). This thesis is not arguing that classical music is in any way inferior, or that its use should be halted, simply that the research should no longer be restricted to this style of music, mainly because it is not the only style of music to which people feel emotional responses. Panksepp's series of experiments (1995) demonstrated this in terms of music eliciting "chills" for the participants. Each participant was asked to bring in music that moved them emotionally, and not one of the seven pieces that evoked the most chills on average was classical.

Many arguments have been made against the use of popular music[1], most of which are unfounded when considered within emotional response studies. First, researchers may say that it lacks the sophistication and history of classical music, and therefore is not worthy of study within the academic realm. If a researcher truly strives to understand the human condition and what stimuli drive people to such fantastic emotional highs and lows, this elitist attitude should be abandoned. In order to analyze the emotional responses that participants experience, they must be presented with stimuli that have the potential to propel them to these states, and classical music does not necessarily catalyze this reaction in everyone.

Second, it is believed that the lyrical content may intrude upon the emotions elicited by the music itself, thereby introducing yet another confounding variable into an already complex topic. For many listeners, the lyrical content acts as a facilitator for the emotions evoked by the music. They do not hear the music as lyrics on top of instrumentation. They hear the song in a complete form, which will not elicit a response if the song is not intact. London (2002) discussed listener preference for lyrical content in songs in comparison to instrumental music by stating that, "This is not because they do not understand symphonies and string quartets, but because perhaps it is those other contexts — where the musical gestures and expressions can be informed by lyrics,

[1] The term "popular music" is used to signify all genres outside of the realm of classical music.

images, and stories — that the music itself can be most meaningful" (p. 36). If the goal of the study is emotivist in nature, the content of the emotional experience is the dependent variable, and only experiments that are surveying the response to specific stimuli should be strictly controlled regarding lyrical content.

Third, some researchers will argue that classical music is more emotionally moving than any other, so it is the clear choice over any type of popular music. Again, not to discredit the emotional impact of classical music, but this belief is about as far from the truth as can be. Anyone who has ever seen the enraptured attendees at a live electronic music concert, the revved-up crowd at a heavy metal show, or the whirling dervishes at a Sufi music gathering have living proof of the emotional heights provided by non-classical music. The differences between a few of these types of music were analyzed in a recent study that set out to determine the effects of techno music and classical music on emotion and neurotransmitters (Gerra, Zaimovic, Franchini, Palladino, Giucastro, Reali, Maestri, Caccavari, Delsignore, & Brambilla, 1998). They found that techno music significantly increased heart rate and systolic blood pressure, as well provoking an increase of neuroendocrine production (β-endorphin, adrenocorticotropic hormone, norepinephrine, growth hormone, and cortisol), while classical music had no significant effects on any of these biological factors. Results such as these continue to bolster the need for studies that will examine the emotional effects of popular music styles.

3.3.2 Familiarity

Familiarity is the next concern for stimulus selection, interacting greatly with the style(s) of music chosen for the experiment. The participants' level of familiarity with the music can greatly affect the power of their emotional response (London, 2002; Gabrielsson, 2002). It is not only important to consider participants' familiarity with the style of music, but also their familiarity with the artist/composer, the album (if applicable), and the song itself. Considering the fact that so many new styles of music are combinations of multiple styles, one artist/composer is not necessarily going to sound altogether similar to another, even if they are listed within the same style or genre. Therefore, if a participant feels a strong emotional reaction in response to the songs of one artist, it is not acceptable to assume that a similar level of emotional response will occur if they listen to an unfamiliar artist from the same genre. Genre classifications work well at a high level of music categorization, but personal musical preferences do not always succumb to accepted genres. Genres change so quickly that using them as a means to determine participant familiarity with the stimuli can lead to false familiarity ratings. For example, McLeod (2001), when talking about the electronic/dance genre, refers to it as "a metagenre that is constantly breaking apart, recombining, and making obsolete numerous subgenres on a yearly basis" (p.73).

Familiarity also must be controlled on a temporal scale, in terms of each participant's last hearing of the stimuli. If a participant listened to one of the stimulus songs even 24 hours before the experiment, it may have a detrimental effect on the potential emotional response. In Grewe et al. (2005), the "chills" recorded by the participants as they listened to the same piece of music each day for 7 days changed significantly in both number and location. The greatest effect occurred after the third audition of the piece, resulting in the remaining four sessions being almost without any "chills" experienced at all. The main difference between the first and second sessions was the change from long, sustained "chills" in the first to short, punctuated "chills" in the second. This is a very characteristic difference that could easily present itself in any study that has not controlled the listening habits of the participants before the experiment.

3.3.3 Length

The third concern is the choice between full songs and excerpts. Excerpts have been used very effectively within cognitivist studies (e.g. Filipic & Bigand, 2004), and also used in a few emotivist studies (e.g. Vieillard, Bigand, Madurell, & Marozeau, 2003). The problem with excerpts in an emotivist study surfaces when they become too short, which ends up effectively turning it into a cognitivist study. For example, in Vieillard et al. (2003), the participants were given a series of speaker icons on a screen, each playing a separate 20-40 second excerpt, and were asked to freely group them in as few or as many groups as they wished. With the correct instruction set (discussed in 3.5), this study had the potential to allow accurate categorization of the participants' actual emotional experience. The span of time for each excerpt was long enough that the participants could let their emotions flow naturally. Yet, in the second experiment, the researchers cut the length of the excerpts down to 1 second each. This extremely brief length of time does not allow for any elicitation of emotion. Taking a look at Scherer & Zentner's formula (2001) for the evocation of emotion from music, the experience involves the complex intersection of "structural features, performance features, listener features, and contextual features". Simply contemplating the first set of features, it is clear that a 1-second excerpt would elicit no emotion. The emotional experience relies heavily on expectancy violation built into the structure of the music (Meyers, 1956; Grewe et al., 2005), and expectancies cannot be set up and subsequently violated within the course of 1 second. The participants would have had to report on either the emotions they thought the music might have elicited had it continued past the span of 1 second, or, very simply, the emotion they felt the music expressed. Supporting this claim, there was a higher correlation between the groupings of the 20-40s experiment and the 1s experiment for musicians than nonmusicians. The researchers attributed this to the musicians' ability to "activate their explicit knowledge about musical structures to complete partial information derived from the 1-s excerpts while nonmusicians base their judgments merely on the available sound" (2003, p. 237). This solidifies the cognitivist aspect of the study, because they support their data with the assertion that the musicians did not actually feel any emotion.

Many considerations are involved in the selection of experimental stimuli, all of which can cause anomalous data if not monitored correctly. Musical genre must first be assessed within the terms of the experimental goals. If the study is meant to examine the emotional response to different periods of European classical music, then selection of classical music is a clear choice, but if the goal of the study is to analyze trends of evoked emotions across the general public, the researcher must do further research into the styles of music that are most often listened to by the public with emotional response in mind. Assumptions cannot be made regarding the emotional evocation capabilities of one genre versus another. Also, music with lyrical content can be used depending on the desired independent variables of the study; if the researcher wishes to analyze the effect of seated versus standing positions on the emotional response, lyrical content is irrelevant as long as the chosen music evokes emotional responses in all the participants. Participant familiarity with the stimuli must be recorded in as great detail as possible to account for any idiosyncratic emotional responses to the stimuli, with attention also paid to the last time each participant has heard the stimuli, if at all. Lastly, the length of the stimuli can affect the ability of the music to actually evoke any emotions at all. In the context of a cognitivist experiment, length should not have much effect on the recognition of expressed emotion, as seen in Vieillard et al. (2003), but care must be taken in emotivist experiments to make sure there is enough length to sufficiently offer the structural qualities that greatly influence evoked emotion.

3.4 Experimental Environment

The effects that a listening environment can have on the emotional response have been consistently overlooked in many experiments. Levels of comfort and emotional disturbance will change along with the environment in which the experiment takes place. The first consideration is the location of the study. Has the participant spent time at this location? More importantly, has the participant chosen to listen to music at this location? Often times, in order to control as many experimental variables as possible, the answer to both of these questions is no. Most experiments take place in a laboratory of some kind, even more often in a soundproof booth. Anyone who has spent time in a soundproof booth understands that they are a bit disorienting. As soon as a participant steps into the booth, they are met with a dead silence most participants outside of the academic world are unlikely to have experienced. A noted difference is experienced just as the threshold is crossed, and this can be jarring for participants. Most participants are used to listening at home, at work, at concerts, on portable music players, and all of those situations feel comfortable enough to allow their emotions to flow freely. This clinical environment resembles none of those situations and can have a detrimental effect on the ability of the participants to experience the emotions they would normally experience when listening to music. Finnäs (2001), in his exhaustive review of empirical research regarding the effects of live music versus pre-recorded audio/visual stimuli, discusses an experiment run by Reinecke in 1980 comparing a live music setting against a laboratory setting. He recalls that, "The live music was more often associated with positive characteristics such as enjoyment, humanity, romance, and softness. However, the same music in the laboratory presentation more often connotated negative feelings of fear and dislike, and musical-acoustical characteristics such as atonality and noise" (p. 61).

McAdams et al. (2004) made a step in the right direction with the experimental setting of the Angel of Death project. By temporarily outfitting two concert halls with the devices needed for 64 participants to record continuous ratings of experienced emotional force, they were able to successfully recontextualize an experiment to allow the participants to remain in a familiar setting. The task at hand (moving a one-dimensional slider as their emotional force changed) was novel to the participants, but it was adopted quickly, with some participants having, "indicated spontaneously in the questionnaires that the focused task enhanced the quality of the listening experience, because it drew them into the piece more" (p. 345). While most experiments may not be able to leave the laboratory setting due to hardware, software, or various other constraints, experiments like this one set an example for future research with emotivist goals, and should influence researchers to attempt an experimental design that would allow the progression towards a more naturalistic setting.

If within a laboratory setting, there are a number of variables that can be controlled that will make the setting more or less conducive to natural emotional responses. One of the main concerns is the physical comfort provided to the participant. If a participant is used to reclining in a plush chair while listening to their favorite music at home, then it will feel awkward and uncomfortable to sit straight up on a folding metal chair. In the same vein, if the participant listens to most music while dancing at a live concert, being forced to stay seated in the same folding chair may be very detrimental to the recorded emotional response. When the goal is to have the participants experience as powerful emotional states as possible, an effort must be made to ensure the comfort of each participant. While there are no standards regarding how to set up an adaptive experimental environment, some forethought is beneficial. For example, if one participant likes to sit in a comfortable chair when they listen to music, and another likes to lay down when they listen to music, there is a simple compromise to accommodate both preferences: have a small couch in the experimental space. Then, most people can be comfortable, and in the rare case that a participant usually sits in a stiff metal chair when listening to music, the researcher can unfold one in front of the couch for that session. Many situations such as these can be accommodated relatively easily. One such situation is the lighting of the experimental space. Some participants may prefer it well lit, while others may ask for it to be pitch black. Recent research has shown that darkness enhances the amplitude of the human startle response (Grillon, Pellowski, Merikangas, & Davis, 1997), which can affect the ANS of a participant in response to certain expectancy violations. Therefore, a dark room may be more conducive to "thrills" than a lit room. In this case, the determination should be made prior to the experiment about whether or not the researcher is testing this hypothesis. However, in an emotivist study that is not testing if one of these variables affects the emotional response, every measure should be taken to make the participant as comfortable as possible. The researcher should express this concern to the participants at the beginning of the experiment, making sure that the participant feels free to request environmental changes. The only caveat is that each of these environmental variables must be recorded if they are modified. The data about the individual environmental conditions has significant potential for later analysis, and may help in the grouping of participants by listening habits.

Another important factor that is an integral part of the experimental environment is the proper management of the equipment used for the experiment (both to control and to amplify the sound). For a large number of experiments, the controlling equipment is comprised of one or more computers with which the participants must interact. Some practical concerns arise when this is the case, such as the possible inadvertent interruptions from the visual stimuli. If the participant is not required to manipulate the computer itself during the audition of the stimuli, it is advisable to darken the monitor, either by cutting the power to it or using a power management setting within the operating system to allow the screen to be darkened by a mouse movement or keyboard command. This will ensure that there are no unnecessary visual stimuli that could potentially take the focus away from the auditory stimuli.

The equipment involved in the stimulus playback can affect the environment as well, mainly related to the choice between headphones or loudspeakers for sound presentation. Crane (2005) surveyed the literature surrounding the use of headphones with personal audio technology, and found that headphones can act as a security blanket, of sorts, providing a "separate experiential reality" that becomes "a mood or experience enhancer that the user manipulates to effect disposition or frame of mind" (p. 11). With this level of connection to the headphone listening modality, forcing a participant, who has developed this bond to headphones, to listen on loudspeakers could negatively affect their ability to respond emotionally. Due to the general lack of research involving evoked emotional responses, no empirical research has been done on the emotional effect of headphones versus loudspeakers in an experimental context, yet it seems the difference may be significant based on this preliminary research into personal affinity for headphones. If the goal of the research is to record the most genuine emotional responses, and does not include the type of sound presentation as an independent variable, the best decision may be to let the participant choose and, as always, record their selection for data analysis.

One of the final concerns for the experimental environment is whether to run a single participant or multiple participants at a time. In some situations (as in a live concert setting), there is no way to avoid the multiple participant scenario and the potential problems that are involved. McAdams et al. mention some of these problems: "Among the disadvantages, one might cite the potential distractibility of listeners caused by the activity on the stage and neighboring listeners" (2004, p. 345). Another disadvantage that should be mentioned when working with multiple participants in clear view of each other's emotional rating device is the potential for group pressure and conformity effects. Asch's studies in the 1950s showed that people in group settings tend to conform to the responses that they see others making (Asch, 1951; 1956). For example, if the audience is using the one-dimensional slider employed in the Angel of Death study and a participant sees out of their peripheral vision that both of the participants sitting next to them move the slider upwards simultaneously, the participant may feel social conformity pressure enticing them to move the slider up themselves even if their emotional force is not rising. The social norms and the need to answer 'correctly' can have a serious impact when participants can see how others are using the rating device. As a simple way to avoid any of the above problems, participants should be tested one at a time, unless the experiment is run in a live setting that negates that possibility, in which case the problems may be mitigated in other creative ways. For example, in the Angel of Death project (McAdams, et al., 2004), the emotional force slider boxes (beige-colored) were interleaved with the familiarity slider boxes (black-colored) throughout the audience to keep neighboring listeners from influencing each other.

With all these possible influences, a researcher must examine the experimental environment thoroughly before any experiment can take place. While controlling a set of these variables according to the goal of the study, e.g. "The effects of lighting and headphone-wearing on the emotional response", may be required, the rest of the variables should be considered, and the decisions recorded, to keep the comfort of the participants as even as possible.

3.5 Instructions

One area of experimental design that can make or break an emotivist or cognitivist experiment is the nature of both written and verbal participant instructions. As discussed earlier, the definitions of expressed emotion and evoked emotion clearly convey their distinctions, but this separation is very often unclear to the potential participants. This lack of clarity requires that the researcher makes the goal of the experiment, and the difference between expressed emotion and evoked emotion, crystal clear to the participant. A few researchers have recently highlighted this distinction and recognized the lack of this methodological concern within much of the current research in the field. Gabrielsson (2002) noted that:

> A pertinent methodological problem is that neither all researchers, nor all subjects clearly observe the distinction between emotion perception and emotion induction. In reading certain reports, one feels uncertain concerning what the subjects reported: their perception of emotion in the music, or their own emotional response, or a mixture of both. (p. 139)

Zentner et al. (2000) tested this theory, and found that the results differed greatly between participants who were asked to rate emotions induced by music and participants asked to rate emotions expressed by music. If the instructions to the participants are not clear enough for a full understanding to be established before the experiment begins, the researcher runs a great risk of collecting useless data. Some studies have taken heed of this warning, and come out with more accurate representations of the participants' evoked emotional responses. One of these was the Angel of Death project (McAdams et al., 2004), which instructed participants very clearly on the handout distributed before the experiment. Some key phrases that clarified the emotivist/cognitivist distinction are: "It is important to remember that your rating should not concern the music itself, but your own emotional response to it," and "There are no right or wrong answers. What matters to us is how you react emotionally to the music." These statements solidify the instructions for the participants, and at the same time, reassure the participants that they should feel free to express their emotions as they experience them.

Nagel et al. (2005), with a similar aim to the Angel of Death study, wrote that, "in contrast to previous studies, participants were instructed to express their own perceived emotions and not to rate the emotional expression intended by the composer" (p. 1). These two examples feature the most important part of the instruction requirements: explaining the negative to reinforce the positive. By the instructions telling the participants what not to rate, the dichotomy between a cognitivist and emotivist study is reinforced.

Many studies have been problematic when their instructional methods are analyzed in detail, causing any number of problems within their datasets. First, consider a series of studies conducted by Madsen along with an assortment of other researchers. The studies began with an attempt to analyze the "perceived aesthetic level" of Puccini's *La Bohème* (Madsen, Brittin, & Capperella-Sheldon, 1993). The instructions given to the participant briefly state that, "This study is an attempt to provide ongoing information concerning what you define as the aesthetic experience. As you listen to the music, move the dial corresponding to your aesthetic response." While this terminology is slightly different than what has been discussed before, Madsen et al. informally equate "aesthetic" responsiveness and a heightened sense of emotion within the first paragraph of the study, so the drive behind the study is clearly emotionally related. Now, if it is analyzed according to the rules being put forth by Gabrielsson and this thesis, its violations are numerous. The category of response was not clarified in the least, even letting the participants use their own definitions of the term that was to be rated. The researchers themselves did not attempt to define the term "aesthetic response."

The next study (Madsen, Byrnes, Capperella-Sheldon, & Brittin, 1993) used the same experimental methodology as the previous one, but sampled an equal number of musicians and nonmusicians, while using five different stimuli across five instances of the experiment. Once again, the terminology was not defined and left completely up to the choice of the participant whether to rate from a cognitivist or emotivist perspective. As before, the instructions asked each participant to rate "your aesthetic response," which suggests that it should be a phenomenon felt by the participants, i.e. emotivist. In the discussion after all five instances, Madsen et al. wrote that, "The aesthetic experience as defined by subjects within these studies (in relation to the Western art music examples), is not an overall encompassing attribute of music that begins with a piece of music and continues in the same way throughout" (p. 187). This statement supports the results of the study being emotivist because the "aesthetic experience" was not found to be an attribute of the music itself, which would denote cognitivism. Yet, even though the overall responses showed an inclination towards emotivist responses, the ambiguity of the instructions resulted in "a high degree of variation across music selections and individual persons" (p. 187).

A few years later, Madsen produced back-to-back studies in 1997 and 1998 using a two-dimensional rating scale, for which the instructions were as follows:

> This study is an attempt to provide ongoing information concerning what you define as exciting/relaxing—beautiful/ugly. You will see that the mouse icon can be moved simultaneously as you listen throughout the music excerpt. As you listen to the music, move the mouse icon on the monitor screen to correspond to your responses of relaxing/exciting and ugly/beautiful. (1997, p. 192)

This instruction set raises many questions concerning the goal of the study. The dimensions may or may not have any correlation whatsoever due to their inherent ambiguities. Taken separately, exciting/relaxing is linked with the arousal scale present on Russell's circumplex model of affect and seems implicitly to be thought of as emotivist in this study, due to the use of the term "your response". The second dimension does not fit in Russell's model, nor does it work alongside of exciting/relaxing if that dimension is taken simply as an emotivist scale. The situation gets more confusing when the "your response" term is linked to this dimension of beautiful/ugly. A person can, theoretically, think of music as beautiful or ugly, but it seems unnatural to consider "your response of beautiful/ugly." Beyond these faults, there seems to be no valid reason for reversing the order of the high/low axes labels the second time they are mentioned. As could be expected, the data for each study turned out very different. The difference cannot be attributed to the stimulus change either (1997: Puccini's *La Bohème*; 1998: Haydn's Symphony no. 104), because it is clear from the graphs of the data (Fig. 2 and Fig. 3) that the rating scale was used very differently for both experiments.

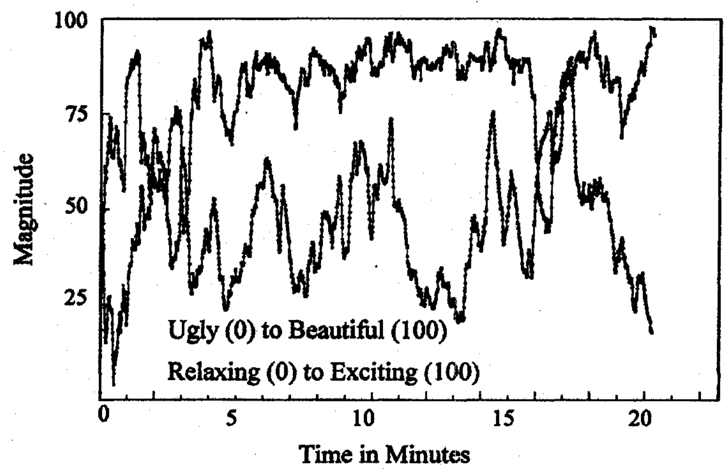

Fig. 2: Responses to *La Boheme,* reproduced from Madsen (1997)

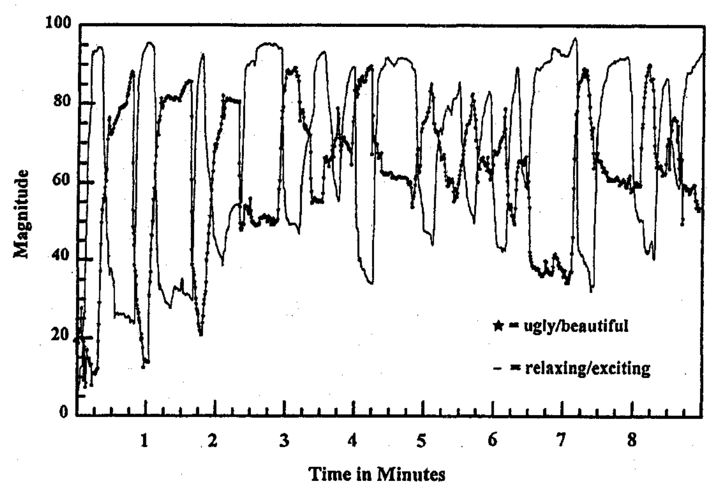

Fig. 3: Responses to *Symphony no. 104,* reproduced from Madsen (1998)

In the 1997 experiment, the correlation between the dimensions was .398, with the exciting/relaxing dimension fluctuating in the middle 50% of the scale as the beautiful/ugly dimension stayed in the top 25% of the scale for the majority of the piece. In the 1998 experiment, with a similar sample (1997: 48 adult music majors; 1998: 50 adult music majors), an identical experimental setting, and identical instructions, the correlation was -.58. The graph of the data bolsters this peculiarity, because the two dimensions fluctuate in opposite directions every 15-30 seconds showing a completely inverse relationship. The data showed that each time the music was rated as exciting, it was also rated as ugly, and vice-versa. According to this, the only configurations of the two dimensions used by the participants are exciting/ugly and relaxing/beautiful.

Clearly, the participants must have been confused by the instructions to the point where they used the rating scale in different ways. Considering that almost every other facet of the experiment was similar if not identical, there are few other explanations for this large discrepancy between rating techniques, especially when the potential misinterpretation of the instructions is a factor.

Instructions may also be present during other parts of the experiment to help ensure that the participants understand fully what they should be rating. In Schubert's doctoral dissertation (1999), he realized this potential problem through some prior research that "demonstrated that participants found it difficult to make the distinction between describing the emotion expressed by the music and reporting their own emotional experience" (p. 252). To remedy this potential problem, Schubert put a written reminder on the software interface that the task was cognitivist.

As is easily seen, the smallest miswording in the instructions can affect the perception that the participant has of the rating scale, mainly because most participants are not familiar with the emotivist/cognitivist discrepancies. Therefore, it is key in either type of study to inform the participants of the emotivist/cognitivist difference and then clearly state which is to be rated.

3.6 Questionnaires

Perhaps the most daunting of the design tasks is the development of questionnaires that not only provide the researcher with the most accurate answers to the research questions at hand, but also ask the correct questions. Questionnaires can have many different roles and placement locations within the experimental process. The three types that are most important for a successful emotivist study are pre-experiment, mood assessment, and post-experiment questionnaires.

3.6.1 Pre-experiment questionnaire

Pre-experiment questionnaires are used to collect the background information needed from the participant for eventual data analysis and, in some cases, data mining applications. The information collected can be split into a series of categories:

- **Demographic/personal:** age, gender, race, handedness, socio-economic stature, education, occupation, and any other information that wouldn't fit into the other categories. The one stipulation from a human research ethics standpoint is that no personally identifying data should be collected.
- **Music education:** complete history of professional and nonprofessional music education.
- **Music playing/writing habits:** status of musical instrument playing, frequency of play, styles played, environment for playing, whether or not the participant composes original music, style of original music composition.
- **Music listening habits:** listening frequency, motivation for listening, listening environment (location, furniture, companionship, etc.), source of music (albums, playlists, etc.), type of playlists (original, random, etc.), equipment for listening (headphones, speakers, etc.), and any other information that could potentially inform the researcher about the way in which the participant listens to music.
- **Musical experience:** favorite genres/styles, favorite artists, level of familiarity with genres/styles, frequency of listening to favorite genres/styles, motivation for listening to different genres/styles, and any other questions that could inform the researcher about how and why the participant selects the music that they do.
- **Experience with the stimuli:** level of familiarity with genre/style(s), artist(s), album(s), and piece(s)/song(s) being used as the stimuli in the current experiment, specific emotions/memories related to song(s) being used as the stimuli in the current experiment.

Each of the above categories is essential for almost every emotivist experiment. For cognitivist experiments, however, the need for in-depth knowledge about the participants' listening habits and experience with every musical style, artist, and song will not be as important in most studies due to the distance from individual musical preference as well as the established universality of expressed emotion recognition. Questionnaires regarding listening habits and musical listening experience are suggested to give more information about how music could affect a participant's emotional state, not about how they may perceive an emotion in the music. Starting from the top, there is the demographic category. This is the most basic of all the sections, contributing the data collected for just about every experiment regardless of topic. Next is the music education section, which is often present in emotional response study questionnaires due to the aforementioned tradition of using musician versus nonmusician as the main sample selection criteria. One must ensure that the questionnaires do not go in-depth to the point that the participant grows weary from the length. The types of questions used in the music education section are relatively commonplace, so there is no need to go into detail about them. The remaining four categories are the novel ones that are not often found in current emotional response studies.

The music playing habits category is designed to extend from the musical education section, and find information regarding the practical side of the education, i.e. "now that the participant's musical education history is understood, what are they doing with that wealth of knowledge?" Through questions in this section, the researcher can learn about the current state of the participant's musical proficiency. What is the benefit of knowing that participant #5 was trained for 12 years on piano if the participant has now stopped playing piano after growing to despise it, due to forced lessons for the duration of their childhood? Current information like this could contribute very valuable data to an emotivist experiment. At least it could explain why the emotional force of participant #5 seemed so uncharacteristic every time the piano line started up again in the stimulus. This data will be largely nominal, with the exception of any information regarding how frequently the participants play music. Free response of musical genre/stlye can lead to problems during data analysis due to the wide possibilities of responses, but should not be overlooked because the specificity of musical genre/style can be very precise. The researcher may need to analyze the responses by hand and find logical groupings of genres/styles prior to inclusion in the study.

The data offered from the category of music listening habits can be beneficial in many ways, even beyond the scope of the experiment. At its surface, the questions provide the researcher with valuable information about how often the participant listens to music, the role of music in the participant's life, and how comfortable each participant is going to be in the experimental environment based on a comparison to their normal listening environment. Besides that, each question helps to further the knowledge about the listening habits of large populations. If this data were collected for every emotional response experiment, it could easily be uploaded by each researcher to an online database, along with the participants' demographics, resulting in a global database analyzing the listening habits of thousands. Some of these questions, such as playlist source, may seem arbitrary, but can truly affect how a person listens to music, thus affecting the emotions they might feel. The data characteristics of this category are similar to those of the music playing habits: mostly nominal, with frequencies recorded as ordinal, and freedom of response required when genre/style is considered.

Musical experience is a category that must not be overlooked when the study has emotivist goals. The emotions people feel in response to music depend so much on the contextual and social associations developed with styles, artists, albums, and songs, therefore the more information the researcher has about these individual connections, the more likely the recorded emotional response could have significance, and the more likely there might be possible groupings amongst the participants. The stimuli experience category is closely linked to the musical experience category, but must be tailored to each individual study in order to find out every bit of information possible regarding the participant's relationship to the musical stimuli. Much free response will be required for the data in this category, with Likert-type scales used to record levels of familiarity and intensity of prior emotional/situational associations the participant may have with the stimuli.

3.6.2 Mood assessment

The next type of questionnaire allows the researcher to determine the current mood of each participant moments before the experiment occurs, which provides a great deal of knowledge about the levels of emotional and mood disturbance and their effects on the magnitude of the participants' emotional responses. This procedure is relatively unique within the setting of an emotivist (or cognitivist) study, but it seems to have the potential to clarify data analysis, and is simple and relatively unobtrusive for the participant. The researcher should not attempt to develop their own questionnaire for this section, mainly because there is a large selection of mood assessment scales available, many with very well established reliability and validity within the field of experimental psychology and psychometric testing methods. The one used in the experiment for this thesis is called the Profile of Mood States, or POMS, and has been used in music research in the past (Davis & Thaut, 1989; Matsuura, 1998; Smith & Noon, 1998). Depending on the goal of the study, this questionnaire can be used multiple times to assess the change in mood state as the experiment progresses, or more specifically, the change after listening to certain stimuli. The researcher must take notice that the questionnaires often take nearly ten minutes to complete, and may themselves negatively affect mood if too many must be filled out over a short period of time.

3.6.3 Post-experiment questionnaire

The remaining type of questionnaire is used to gather participant impressions about the stimuli, the equipment used for the emotional rating, and various features about the impact of the experimental setting. The arrangement of these questionnaires is up to the researcher; there can be one all-encompassing questionnaire at the end of the experiment, one after every stimulus, one after each use of a rating scale, etc. The importance within these forms is the use of rating techniques that do not force the participant into a false response because of the participant's desire to 'answer correctly'. As Orne (1962) wrote, "as far as the subject is able, he will behave in an experimental context in a manner designed to play the role of a 'good subject' or, in other words, to validate the experimental hypothesis" (p. 778).

There are many instances of post-experiment questionnaires that could offer the demand characteristics needed by a participant to affect their responses in this way within the series of studies conducted by Madsen that were analyzed in section 3.5. After the stimulus was presented in Byrnes, Capperella-Sheldon, and Brittin (1993), in which the one dimensional controller was used by the participants to rate their "aesthetic response," the participants were asked, "Do you feel the movement of the dial corresponded to that experience?" After this question, the participant is given the choice of "yes" or "no".

This lack of options forced the participants to choose a side with no gray areas possible. Thinking logically about this, it is easy to see that most participants are going to answer in the affirmative due mainly to pressure induced by perception of the demand characteristics of the study. They are not able to rate how correctly the dial movement corresponds on a scale from one to ten, thus they could easily have felt that the entire study would be a waste if they flat out answered "no." The data could have been much more valuable in determining the validity of the controller if the question had been asked on a Likert-type scale from one to ten.

Another type of question found in this study, and a few others, is difficult to answer with any degree of reliability. The question asked was, "What was the highest magnitude of this experience compared to others you have had?" For this, the researchers used a one to ten Likert-type scale, but the damage was already done with the question. This question assumes that every participant can not only remember every other emotional or "aesthetic" response they have ever had, but also that they would be able to reliably compare the one they just experienced to a great abundance of prior experiences. A question such as this is too daunting to be so quaintly asked, considering that quantification of all past experiences in relation to the current experience will result in erroneous comparisons. Winkielman, Knauper, and Schwarz (1998) discovered this in a study related to the self-report of concurrent and retrospective analyses of instances of anger: "Self-reports pertaining to different reference periods cannot be directly compared with regard to either the frequency or the intensity of the event" (p. 726).

A similar problem was found within Sloboda's (1991) exploratory study that attempted to learn about the intricacies of physiological responses to music. The study gave the participants a list of physiological effects, e.g. shivers down the spine, laughter, tears, etc., and asked them to "nominate up to three pieces of music in which they could remember experiencing one or more of these physical responses within the previous five years." The participants were then asked, for each piece, to "identify the nature of the response, its consistency, and the proportion of listenings on which the experience occurred." This is a much clearer, and more thought-out approach than that in the Madsen study, yet it still has its problems. Identification of the pieces of music is relatively simple, as is the selection of effects from a predetermined group of possible ones. The difficulty is found in the determination of the consistency of the specific effects and the proportions of the responses. Most people do not keep mental notes of the percentage of times that listening to a song causes a physiological reaction. They could easily remember that a song causes a reaction, and even whether it causes the reaction somewhat frequently, but asking to quantize these vague memories has potential to be problematic, especially when considering the time scale distortion evidenced in Winkielman et al. (1998).

Questionnaires such as all three types discussed can be extremely beneficial, if not necessary for valid data analysis. If all of these participant-related factors are taken into consideration there is a greater chance of finding correlations between participants in similar groups. For example, Nagel et al. (2005) stated that they found, "no inter-individually consistent rules for the relationship between self report and physiological data," but noticed, "smaller groups of subjects that show similarities in their reactions to music" (p. 1). Their participant pool consisted of 35 participants, using stimuli consisting of "pre-selected standard music pieces" and "[the participants'] favourite pieces." After each piece, the participants had to answer questionnaires, "concerning their associations related to the respective piece of music." Also, the post-experiment questionnaire collected, "information on musical expertise and personality factors." These extra data sets enabled the researchers to analyze the smaller groups that displayed similar responses in the self-report and physiological measures. If they had also collected information on the current mood states of the participants, as well as measures of cognitive ability, they would have had more factors to describe, define, and discretize the groups.

To clarify the recommendations in this chapter, Table 1 is included as a reference. Table 1 is divided into emotivist and cognitivist categories to clarify what needs to be considered within each experimental design category depending on the type of experiment. With the framework regarding the design and procedure for emotivist and cognitivist experiments established, the specifics of the hardware that will record the participants' emotional responses must be examined. In order to develop the best possible interfaces for the continuous response, an in-depth review of previous controllers is presented, followed by the design specifications of the controllers that were constructed for this experiment.

Section	Emotivist	Cognitivist
Participant Selection	• Consider participant selection criteria beyond musician/nonmusician. • Avoid assumptions about musical background based solely on musical education.	
Stimulus Selection		
– Style/Genre	• Never underestimate a style/genre's ability to evoke an emotional response • Lyrical content should not be altogether avoided, and allowance in an experiment should be assessed on a case-by-case basis.	• Researchers should begin to experiment with genres that do not have clearly expressed emotions
– Familiarity	• Familiarity with stimuli must be determined from the genre all the way down to the specific piece. • The participant's should not have recently listened to any of the stimuli to ensure maximum emotional response.	• Style/genre familiarity should be sufficient to assess a participant's ability to perceive expressed emotions in stimuli.
– Length	• Full songs are preferred over excerpts. • All excerpts must be long enough to convey structural features integral for emotional evocation.	• Length is inconsequential, considering the findings by Vieillard et al. (2004)
Experimental Setting	• Experiments should be run in a natural, yet controlled, environment if possible. • User-specific modifications of the experimental setting may be beneficial in maximizing emotional experience, as long as modifications don't interfere with independent variables.	• Comfort level in experiment should not affect responses, so strict environmental variables should remain constant.
Instructions	• Phrasing of instructions must be precise and explicit concerning distinction between emotivist and cognitivist studies including the chosen type for current experiment.	
Questionnaires		
– Pre-Experiment	• Questions must extend beyond music education into the realm of music listening habits, music playing habits, and music selection. • Full experience with current experimental stimuli must also be collected.	• Beyond music education, experience with current experimental stimuli should be collected.
– Mood Assessment	• Use pre-existing mood scale to ensure reliability and validity. • Consider running multiple mood assessments during experiment to analyze how mood changes due to listening.	• Mood assessment not required.
– Post-Experiment	• Keep demand characteristics vague to avoid conveying experimental hypothesis to participants. • Avoid the combination of different time scales in emotional memory tasks (e.g. year vs. week vs. current).	

Table 1: Experimental Design Considerations

4. Interface Design Considerations

4.1 Preliminary Research

The impetus for the main experiment in this thesis was born from a pilot experiment run nine months earlier. The experiment was designed to allow a group of electro-acoustic music enthusiasts to rate their emotional force changes on the sliders used in the Angel of Death project (McAdams et al., 2004) while they listened to relatively unknown pieces of music that would be classified within the Noise/Experimental genre. The main goal was originally to examine any patterns of emotional response across the participant group that could be attributed to timbre changes, with a concurrent goal of collecting comments concerning the use of the slider device.

4.2 Pilot Experiment Methods

The experiment took place in the Real-Time Multimodal Laboratory in the Music Technology Area at McGill University on November 9th, 2005.

4.2.1 Participants

Nine music technology students participated in the experiment, with no compensation offered in exchange for their participation. Three of the participants were undergraduates in music technology, with the remaining six studying at the graduate level. The only prerequisite for participation was a familiarity with electro-acoustic music.

4.2.2 Stimuli

The stimuli were selected to be relatively novel to the participants. As the participants and their full music listening histories were unavailable at the time of the experimental design, pieces were chosen based on their obscurity within the genre. The pieces were (in order of presentation): *Happy Audio* by Fennesz (2001), *Munchen* by Masami Akita aka Merzbow (1998), *Espereptic* by Doug Van Nort (2004), *Lens Test* by Kim Cascone (2002), and *A Microsound Fairytale* by Stephan Mathieu (2003). Another attribute sought by the researcher was the lack of any contextual clues within the music that could influence the evoked emotion of the participant by introducing perceptible stereotypical emotions. The stimuli were without discretized melody or harmony for this reason. Lastly, in the absence of melody and harmony, rich timbre changes were sought to provide a large palette of sonic textures from which the participants would react emotionally.

4.2.3 Equipment

The slider boxes were chosen from the actual devices used in the emotional force part of the Angel of Death project. The boxes contained potentiometers that allowed continuous variation of a voltage between 0 and +5 V. They were attached using XLR connectors to two concentrator boxes that routed the signals to an AtoMIC Pro A/D converter through DB25 connectors. The voltage changes for each slider were mapped to MIDI control messages on a 0-127 scale, and the slider number was mapped to the controller number. The data were sent into an Emagic Unitor 8 MIDI concentrator, finally routing them to a Macintosh G5 in which the messages were recorded in Logic Audio Platinum simultaneously with the stereo audio track that was being played.

4.2.4 Procedure

The participants were asked to seat themselves in any of the seats provided in the center of the room. The researcher then gave each participant a questionnaire to be used both during the experiment as well as after the conclusion of the experiment. A very brief explanation of the goals of the experiment was given to the participants, including an explanation of the slider system, and how to use it during the experiment. The use of the sliders was explained in an identical fashion to that in the Angel of Death project. The listeners were asked to rate the continuous force of their emotions, regardless of the kind of emotion they were experiencing. Also, great emphasis was placed on the rating of their evoked emotions, and not those perceived as being conveyed by the music.

The participants were seated in the center area of the room, facing two studio monitors. The researcher sat at a desk behind the participants and controlled the recording of the data and playback of the music with a computer on the desk. As per the request of the majority of the participants, the lights in the laboratory were turned off, leaving only a slight glow from the computer screen that faced the researcher as the only remaining illumination.

Each piece was presented individually, with a break in between each, at which point the researcher would turn on the overhead lights and ask the participants to rate their overall evoked emotional valence for the piece on the questionnaire (positive / negative). After all five pieces, the participants were asked to complete the rest of the questionnaire. This questionnaire is reproduced in Appendix A.1.

After the experiment was over, an informal session was held to discuss how people used the slider, as well as more in-depth elaboration of the comments given on the questionnaires.

4.3 Pilot Experiment Results and Discussion

Besides the interesting information concerning emotional response and timbre that the study yielded, which will not be discussed here, the experiment's secondary purpose helped to highlight the flaws possible in all the aspects of experimental design discussed in Chapter 3. It also supplied a basis for the further research direction explored in this thesis. The main reflections from this study focused on stimulus selection, questionnaire design, and controller design.

Stimulus selection in this study could have been done in a much more rigorous and controlled manner. The fact that a third of the participants were familiar with one or more of the stimuli may have been a serious confound in the data, especially when the main goal was to let the participants' emotion flow without any preconceived notions about the stimuli. Even if prior knowledge of the participants' musical listening history is unavailable, the researcher must be able to account for these effects by determining individual levels of familiarity with the stimuli on the questionnaire(s). It is best to determine this familiarity factor regardless of how controlled subject and stimuli selection is, but it is a necessity if these first two factors are not completely controlled.

The questionnaire used in this study made the first attempt to learn about this specific information, but had been designed with an encompassing knowledge in mind, without drilling down to the levels that are necessary for clear data analysis. For example, the questionnaire asked how familiar the participant was with the styles of music presented in the experiment (rated on a 3-point scale: very familiar / moderately familiar / unfamiliar). While a more finely graded scale, such as a 7- or 10-point Likert-type scale, would have been more appropriate for the question, this is not the most significant problem with the question. The problem lies in the possible misinterpretation. First, the question does not define which "style" the researcher is attempting to analyze. With the melting pot of musical genres that subsumes any attempt at categorization of contemporary music, the genre of the stimuli could easily be misinterpreted. Some would call *Happy Audio* "ambient music" because of its floating, thick texture, while others would call it "glitch music" because the textures themselves are composed of thousands and thousands of granular blips, clicks, and beeps. Clarity is the best practice in every aspect of questionnaire design if any comparison of data between studies is going to take place. Secondly, the question asked only about the genre, and not about the individual artists, albums, or pieces. While this information is not always going to be absolutely necessary, it has potential to be the linchpin of a study; therefore it is always worth the extra tens of seconds that it will take the participant to fill it out.

The controller conundrum that presented itself was discovered during the analysis of the answers to the last question ("How easy was it to control the slider in such a way that your emotional force was recorded transparently?") and during the discussion session after the experiment. Four of the participants wrote that the slider was "difficult" and they "did not have any reference" referring to the fact that they had to look often at the slider to determine their position relative to the absolute limits. When this concern was brought up in the discussion session by one of the participants, all of the other participants nodded in agreement that the scale felt limiting and distracting from the emotional experience. This experiment demonstrated that the design of a controller must allow the participant to monitor and record their experience while minimizing the separation between those two aspects of the directive, i.e. unifying the experience of the emotional response with the physical act of recording it on the controller. Many controllers in the past have potentially been susceptible to this problem, so, in order to attain a full understanding of the features that can make or break a continuous response controller, all previous instances of this type of device needed to be analyzed.

4.4 Early Continuous Response Interfaces

One of the first studies to use continuous responses to study emotion was performed by Clynes using his sentograph, developed in 1977. This device was activated with the middle finger and was capable of measuring both downward pressure and lateral pressure in the one dimension away from and towards the participant. Clynes asked the participants to press the device while they imagined certain emotions, resulting in patterns in the response curves. These patterns surfaced in the majority of the participant responses as very distinct for each type of emotion imagined. Clynes described them as "emotional fingerprints," being natural representations of our physiological responses to emotions, e.g. pressing harder when angry links to higher physiological arousal due to that emotion. Although Clynes did not use his device during the act of listening to music, his study was one of the first to demonstrate the potential of studying emotion on a temporal basis. De Vries (1991) took the sentograph and brought it into the realm of a music listening experiments, demonstrating its beneficial use as a controller for measuring the change in emotional response. The participants were not aware that the researchers were studying emotion, as they were instructed to tap the sentograph button along with the beat of the music. This method differed from Clynes research, in which no deception was involved, and could have a significant effect on the data. If this controller were to be used again, it would be very useful to test it with the same experiment, only informing the participants about the goal and how to consciously use the sentograph. De Vries (1991) also found that the participants had very idiosyncratic ways of using the sentograph, offering the explanation that "their general way of pushing reflects their emotional disposition at the time they enter the laboratory" (p. 61). This statement supports the proposed use of pre-experiment mood assessment scales with the intention of determining the emotional state of the participant. It also indicates that the emotional state while entering the experiment may not only effect the potential emotionality of the participant, but also may dictate the way in which a participant will use the supplied controller.

Goldstein (1980) made the next step in continuous response with his pioneering research into the tracking of "thrills". While no controller in any conventional sense was developed for this study, Goldstein utilized the dexterity and familiarity of finger manipulation to ease and simplify encoding. As the participants listened to music, they held up a finger every time they experienced any kind of "thrill" (or physiological response), and kept it raised for the duration of the experience. Each finger denoted a different intensity of the "thrill," allowing for a dimension of the experience besides duration to be recorded. This research was the first to study evoked "thrills" and their progression over time, but it had one significant problem: the cognition required to quantize the intensity of the experience and select one out of five levels (fingers) can easily delay the actual response timing, thus skewing the data. Also, this combination of cognitive load and successive motor response can interrupt the experience of the thrill itself, either shortening or deadening it altogether. Gabrielsson (2002) supports this by noting that "Several studies…indicate that emotional responses and intellectual (objective, analytical) responses tend not to occur together" (p. 124). With this in mind, the need to make an analytical judgment of the intensity of the response on a discrete scale may interfere with the actual response. Considering that intensity is a very important, and dynamic, aspect of "thrills," this problem might be avoided if the controller used an analog sensor of some kind (strain gauge, pressure sensitive resistor, etc.) to record the intensity changes.

Another early device that was based on "organic" processes similar to those in Goldstein's experiment was the "tongs" designed by Nielsen (1983). The simple controller consisted of a pair of spring-loaded metal tongs with a potentiometer measuring the amount that they were squeezed. Nielsen did not develop them to be used to measure emotion per se, but to allow continuous ratings of "musical tension." Nielsen believed manipulation of the tongs to be gesturally relevant when considering changes in perceived musical tension. Unlike the Goldstein study, which had the participants only make judgments when a "thrill" occurred, this study was the first to require the participants to make judgments continuously throughout the span of a piece of music.

From a design point-of-view, the gestural relevance of the squeezing action is helpful for the acceptance and ease of the device for the participant, but may become an annoyance when the stimulus is long in duration due to the constant muscle activation required for the experiment.

4.5 One-Dimensional Digital Continuous Response Interfaces

With the increasing power and decreasing cost of computers, the next step in continuous response controller design was the introduction of digitized systems. Gregory (1989) made some of the greatest progress in this direction with her design of the Continuous Response Digital Interface (CRDI), developed at Florida State University (Fig. 4).

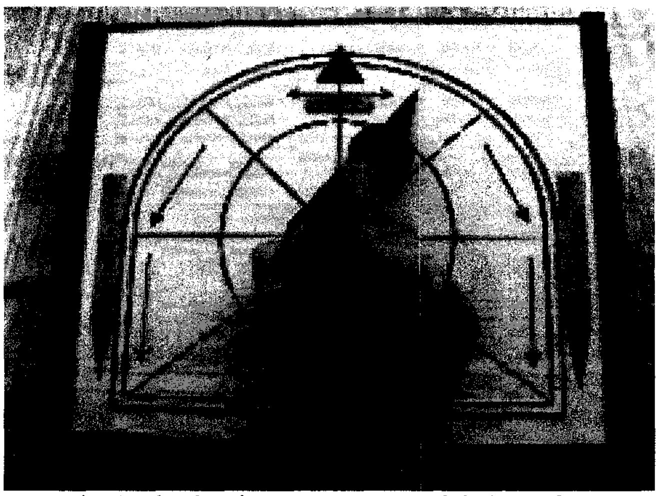
Fig. 4: The Continuous Response Digital Interface

The controller consists of a user-manipulated dial-style potentiometer with a range of 256 degrees. The voltage changes registered on the potentiometer are digitized with an 8-bit A/D converter and recorded on a computer. Underneath the dial is an overlay that can be modified to suit the experiment, providing more flexibility for the controller. This flexibility was exploited within the wide range of experimental situations in which the device was used. Soon after its inception, it was used in focus-of-attention studies (Capperella, 1989; Madsen & Geringer, 1990) with an overlay that showed different instrument groups, asking the participants to choose to which group they were currently listening throughout a piece. It was also used in musical preference studies (Brittin, 1991; Madsen, Capperella, & Johnson, 1991) utilizing various configurations, including happy and sad faces on the overlays. Brittin (1991) was the first to attempt a two-dimensional configuration with two CRDIs that were to be continuously moved during the course of the stimulus (one rated level of musical preference and the other allowed categorization of genre).

Then, in 1993, Madsen and Frederickson ran an experiment using the CRDI to replicate Nielsen's 1983 study of perceived musical tension. A pilot study gave the researchers a design for a "tension" overlay (Fig. 5) that they believed would correspond to perceived musical tension:

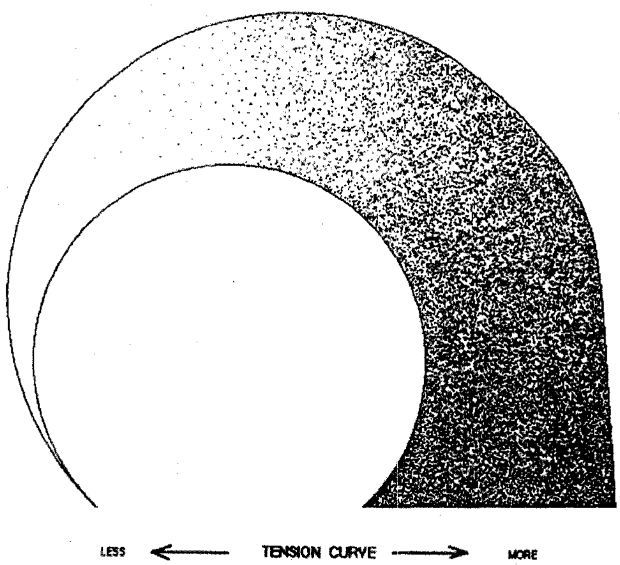

Fig. 5: CRDI musical tension overlay, reproduced from Madsen & Frederickson (1993).

While their results yielded response curves that roughly corresponded to the original graphs in Nielsen's study, some of the conclusions made about the use of the controller seem very misinformed. Madsen & Frederickson claim at the end of the article that the "replication of tension perception using a different measuring device combined with the results of the pilot study, indicate that certain visual patterns are quite different in suggesting 'tension'" (p. 61). This claim is unfounded because a study that looked at shaded images and rated them in terms of *visual* tension does not prove beyond a reasonable doubt that the participants even used the visual aid in rating *musical* tension. Considering that the response curves were similar to the Nielsen study, which had no visual feedback on the controller, the participants may have simply used the upper and lower bounds of the device to gauge the response intensity. If the visual tension overlay were so valid, the results between musicians and non-musicians would have been relatively close, when in actuality the magnitude of the non-musicians' responses were significantly greater than that of the musicians. Does musical training lessen the amount of black and white shading attributed to high musical tension?

With this preliminary research done into musical tension, Madsen and a group of colleagues produced a number of other studies regarding continuous ratings of "aesthetic response." It is within these studies that the flaws in the overlay design became even more evident. Upon labeling the two sides of the CRDI with "negative" and "positive" (Madsen, Brittin, & Capperella-Sheldon, 1993), the results showed that "83% of the participants did not use the entire range of the dial, notably avoiding the lower end of the continuum" (p. 65). This phenomenon was attributed to the assumption that participants did not feel comfortable judging an "aesthetic response" as negative, due to the unfavorable nature of the term "negative." Verbal responses suggested potential changes to the device, with some participants stating that, "they would have preferred a completely positive dial as they had no negative or completely nonaesthetic experiences" (p. 66). This problem highlights the importance of word choice within every aspect of experimental design. This experiment also reported that the participants, "on occasion wanted to move the dial farther toward the positive side [than the controller would physically allow]" (p. 66). This preliminary feedback already supports one of the major concerns that arose in the pilot study discussed in this thesis: the upper bound of a controller employing a bounded scale can limit the reliable recording of the emotional experience.

4.6 Two-Dimensional Digital Continuous Response Interfaces

With the advent of more sophisticated and customizable graphical user interfaces, the continuous response controller was extended to employ more dimensions with less tactile control. Concurrently in 1996, Tyler and Schubert independently developed two-dimensional continuous response systems that used a mouse to control a cursor within an x-y plane on a computer monitor. The systems were based on Russell's circumplex model of affect (discussed in Chapter 2), with the x-axis representing emotional valence and the y-axis representing emotional arousal. Both researchers had cognitivist goals in mind for the experiments (although Schubert was the only researcher who explicitly defined the difference and kept the wording clear and strict throughout the research). Also, both researchers obtained results with high reliability and validity within their experimental constraints. The system design showed much promise, integrating one of the most widely used models of emotion with a relatively simple user interface. The complexity of this system configuration, however, creates a serious need for design re-evaluation when the goals are not cognitivist.

The first problem is the visual interaction required for the use of the system. Since there is no tactile feedback to keep the user within an axis or to let them know how close they are to the limits, the subject is forced to stare at the cursor on the screen over the duration of the experiment. In a cognitivist study, this would not pose much of a problem, considering that the participants are not trying to experience and quantify their evoked emotions; they are simply recognizing certain emotions in the music, which is a task familiar to most people. In addition, the perception, recognition, and categorization of these emotions are all cognitive tasks. Thus the mind does not have to shift from an introspective mode to the encoding mode required for the task to be done on the device.

The next problem is the use of the coordinate plane in which the mouse is moved. As discussed in Chapter 2, the ineffability of the musical emotional experience often tests the boundaries of a model based on everyday emotions. Therefore, the obligation placed on the participant to categorize their evoked emotional response within the constraints of a model that is not suited to encompass such experiences causes not only a high probability of false assessment, but also an interruption of the emotional experience by the cognitive processes required to find a suitable position in the x-y coordinate space.

Two-dimensional continuous response systems such as these have been used in a few emotivist studies (Madsen 1997, Madsen 1998, Nagel et al. 2005) with varying results. The two studies run by Madsen (discussed in Chapter 3) used this design, based on Tyler's research (1996), and produced varying results that indicated differing controller use amongst the same experimental methods. Nagel et al. (2005) used Schubert's (1999) research to develop software called "EMuJoy." This system was identical to the previous, but it also allowed for "chill experiences" to be recorded with a click of the mouse button. Through the study, they found very little similarities between participants' use of the two-dimensions, yet found a high correlation between the reported chill responses and a number of physiological responses that were recorded during the experiment (skin conductance response (SCR) and heart rate).

4.7 New Interface Designs

After a survey of the designs implemented for continuous response systems in the past, it is important to determine the features of a new system that might overcome some of the downfalls of prior designs, as well as being optimized for recording of evoked (as opposed to perceived) emotions.

4.7.1 Transparent Emotional Encoding

One of the main goals in response device design is transparency of emotional response encoding. This feature ideally allows the experience of the evoked emotions within the participant and the subsequent recording of the response to occur with as little interference as possible. Considering this goal, a system that requires the participants to have visual interaction with the device is unacceptable. Breaking down the process makes this necessity clear: When a person experiences an emotional response, their natural tendency is not to look at a screen and think to themselves, "How far did I just move the cursor during that experience? Did I move it enough? Was the amount I moved it more than the last time I felt the emotional experience? How close is the cursor to the top limit of the device? Should I move it back down because all these questions I'm asking myself are taking me out of my powerfully ecstatic state and making me a little nervous?" A much more natural, and transparent recording of the response would be for the subject to squeeze an object, such as an armrest or a spring. This is the body's natural link from the autonomic nervous system in response to emotional changes. The action that the participant must take to record their experience must be 'gesturally relevant,' meaning that the action, including its direction, amount of pressure, and proprioceptive feedback, must be natural extensions of the emotional feeling itself. The new interfaces in this thesis were designed to test different gesturally relevant movements and analyze their effectiveness.

4.7.2 Interface Size

The size of the interface can have a great impact on how and where it is used. Ergonomics must be kept in mind in order to make the experience as comfortable as possible for the participants, which in turn helps to let the emotional response flow unimpeded. If the device is to be held in the hand, it must be of a suitable design for any size of hand (although a different controller will most likely be needed for children), as well as being comfortable for both right- and left-handed participants. The shape of the device should be familiar to the participants in order to increase the overall proficiency in its handling. Large muscle movements must be avoided during the experiment due to the possibility of distracting the experience as it occurs, so smaller devices are usually preferable to larger devices. This problem can be present in many ways, such as needing to move the upper arm to resituate the forearm so that the mouse can be moved further in one direction. The movement would require an interruption of the experience to make the required movements possible. As for location considerations, a smaller size makes the unit more portable, as well as more adaptable to environments outside of the laboratory, such as concert halls or even participants' homes. This is a key factor when considering that emotional responses in and out of the laboratory setting must be studied and analyzed for any attenuative effects that the environment may have. Also, if the effect of physical activity on emotional response to music is desired, which preliminary research shows may be significant (Dibben, 2004), the device must be small enough to be used during the required physical activity.

4.7.3 Rating Scale Attributes

Due to potential problems with emotional interruption when two-dimensional x-y coordinate systems are implemented in the interface, a single rating dimension should be employed for emotivist devices. This does not exclude the use of other ratings that do not occur continuously throughout the experiment, e.g. chill response recordings. These extra responses may be considered as a separate dimension, but they do not require the level of divided attention needed by the x-y plane configuration. One continuous dimension provides great flexibility in the labeling of the rating scale as well.

The other aspect of the rating scale (or more importantly the hardware used to record the scale) is the selection of an absolute scale or an unbounded scale. All the systems developed so far have been absolute in nature, i.e. they have upper and lower boundaries. As found from the pilot study, this limitation can have a negative effect on the response by forcing the participant to visually check their position on the device relative to the limits. Once they perform this action, it may interrupt their flow of emotions and consequently cause a drop in emotional force. At the very least, the participants could use the top boundary in different manners. For example, some participants may approach the top boundary and let the controller hit the maximum, even though their response continues to rise, while others may monitor their position and exponentially reduce their upward controller motion in response to their upward emotional change to always leave a buffer near the top in the case of a stronger response. Since there is no way of inherently knowing which method was employed by the participants (and there is little practicality in explaining the difference and questioning each participant individually), a novel approach is to use a scale that has no boundaries. A second potential problem with the upper boundary is the rescaling it forces when a participant experiences a higher response when they are already at the maximum level on the scale. When this occurs, the participant will, most likely, mentally re-associate this new emotional high with the upper bound of the controller, thus rendering the scaling of the responses for the remaining section of the piece of music to be different than that of the previous section.

Unbounded scales allow for great flexibility if combined with the right instructions for the use of the device. Absolute scales can force the participant to modify the way that they use the device in relation to their emotional experience because of the need to keep their ratings within the pre-established bounds. This problem will never occur with unbounded scales, thus eliminating the need for the participant to habituate the emotional intensities experienced by a stimulus to keep their ratings in the correct locations in the scale. Unbounded scales can be implemented in various hardware configurations, and are the basis for the controllers designed in this thesis. One potential problem with this new type of scale lies in the data analysis. The use of an unbounded scale can allow somewhat idiosyncratic uses of the scale, due to the nature in which the responses must be recorded. For instance, one use of an unbounded scale is to ask the participant to press a button every time they feel an upward movement in emotional force and hold it for the duration of the change. In this case, the researcher must analyze not only the times when the button is pressed, but also the duration it is pressed. However, despite this data complexity, the merits of the unbounded scale provide the flexibility needed to accommodate the wide ranges of emotional changes that characterize human response to music.

5. Main Experiment Design

5.1 Hardware Design

While keeping all the factors of interface design discussed in Chapter 4 in mind, two new interfaces were designed for use in this thesis. The controller was built using the body of a pre-existing USB mouse, called the *USB Grip Mouse*, manufactured by Nexxtech, a subsidiary of Orbyx Electronics.

Fig. 6: Joystick Controller

The body offers a comfortable grip for most hand sizes, as well as full ambidextrous support. The original device offers a trackball under the thumb, two buttons above the trackball, a trigger attached to a micro switch, and full USB support. For the purposes of the experiment, the trackball was removed and the two thumb buttons were secured to the body with epoxy because they would not be used in the study. The trigger was left with the intention of having the participant record "thrills" on this button. With a controller housing selected, the next objective was to find an appropriate transducer for the emotional force ratings.

Many types of buttons, switches, sliders, and potentiometers were considered to take the place of the trackball underneath the thumb. In the end, two were chosen to test differing implementations of the unbounded scale. The first, as seen in Figure 6, was a spring-loaded joystick taken from a gaming controller (while a joystick is not physically unbounded, the directions in the experimental procedure inform the participant to use it as such to rate *changes* in emotional force, not to use its mechanically-bound range to keep track of their *current level* of emotional force). Other hardware could facilitate a rating scale such as this, like a scroll wheel on a mouse, but the Joystick was selected for familiarity and ease of installation. The joystick was mechanically restricted to its y-axis to support the one-dimensionality of the rating scale. It was then installed on the circuit board in place of the trackball and the height of the stick was reduced for comfort. The participant can hold the Joystick in either hand and manipulate the y-axis of the stick with their thumb. The Joystick controller was designed for the participant to rate their changes in emotional force both in upward and downward directions.

The second transducer chosen was a simple plastic button removed from the same gaming controller. Initially, the intention was to use the button in its binary configuration, allowing the participant to press the button when their emotional change would rise and to release it when it would fall. However, a force-sensitive resistor was selected instead. So, the button was installed in the second housing in place of the track ball by affixing a half-inch force-sensitive resistor to the circuit board, with the rubberized contact of the button and the button itself on top. Just as with the Joystick, the participant can hold the Button device (Fig. 7) in either hand and press the button with their thumb. The motivation for the design of this Button controller was to instruct the participants to press the button as if it were binary, then determine later if they pressed harder or softer when the response was more intense. This idea stems from the research of Clynes (1977) and DeVries (1991), which found that the time course of button pressure can be linked to different emotional states.

Fig. 7: Button Controller

Each of the controllers was fitted with a DB25 9-pin male connector connected to the AVR-HID, which is an 8-channel USB sensor A/D interface designed by Mark

Marshall in the Music Technology Area of McGill University. This device is based on the ATMEL Atmega16 microprocessor and is very inexpensive (≈ 20$CAN). The device connects to the transducers with three leads for each one: +5V, signal, and ground. The unit was designed for use within the Max/MSP software environment used for the experiment. The incoming USB messages were brought into the patch using the "hi" object that provides data exchange with USB-based human interface devices. The sample rate for the "hi" object was set at 10Hz. This sample rate is sufficient when considering that a human could not record emotional fluctuations faster than 100 times per second. A slower rate may also be useful in larger experiments if data reduction is important, and will not effect the experiment as long as the rate is fast enough to detect all emotional fluctuations, such as 100Hz.

5.2 Experiment Methodology

The experiment took place in the Music Perception and Cognition Laboratory in the Music Technology Area at McGill University between July 20th, 2006 and July 28th, 2006. The listening portion of the experiment was located within an Industrial Acoustics Company 120act-3 soundproof booths located in the lab.

5.2.1 Participants

Twenty-four participants took part in the experiment. They received $10 in exchange for their participation. The participant pool was made up of 19 males and 5 females, aged from 20 to 41 years (Mean (M) = 26, Standard Deviation (SD) = 5). Twenty of the participants were right-handed. Twenty claimed to have had music lessons starting at a mean age of 7 (SD = 3), with 18 of those having private lessons.

5.2.2 Stimuli

Stimuli were selected by each participant. They were asked to bring three pieces of music for which they felt a strong emotional response relatively consistently and with which they were moderately to very familiar. The distinction was made with each participant prior to the experiment that they were not to bring in pieces in which they could recognize a lot of emotions, but one that evoked strong emotions in them. The three pieces could be of any genre, style, era, length (within reason), and could contain or not. The only requirement was the evocation of strong emotional responses. These guidelines were selected because the purpose of the experiment was to analyze the participants' preference of controller and the controller's ability to track emotional changes over time. No control stimuli were needed since the experiment did not aim to determine how music affects emotion, but simply what is the best way to record the emotional changes as the music is heard. The stimuli were brought to the experimental session on a CD or flash memory, or could be located on a web server. The files were then moved to the experimental computer. To randomize the order of presentation, the researcher chose the order based on the alphabetical order of the three stimuli.

5.2.3 Equipment

The experiment was conducted on Max/MSP installed on an Apple G5. The audio was sent from the computer to a Grace m904 High Fidelity Monitor System, which routed the sound to two Dynaudio BM6A active nearfield monitors. The computer screen, monitor system, and monitors were located on a desk at the back of the booth with a rolling, adjustable, upholstered desk chair placed in front of them.

The AVR-HID converted all the signal voltages from the three controllers to a range of 0-1024. The available controllers will be referred to as "Joystick", "Button", and "Slider". Joystick and Button were developed specifically for the experiment, and Slider was modified from the original design in McAdams et al. (2004) by removing the rubber bumper that denoted the "Don't know" zone and making the connections compatible with the AVR-HID device. The Joystick was centered at a value around 512, with a downward push bringing it down to 0 and an upward push bringing it up to 1024. The Button had a value of 1024 at its resting state, with very hard forces bringing it down to about 200. Both Joystick and Button had the "thrill" button located in the trigger, which would produce a value of 0 if not pressed and a value of 50 if pressed. The Slider's boundaries were mapped to values of 0 for the "weak" label and 1024 for the "strong" label. These values were recorded throughout the listening session using the *coll* object in Max/MSP, then exported to text, formatted using Sed (a freeware command-line program for running batch edits on a text file), and imported to Matlab and SPSS for analysis.

5.2.4 Procedure

The participant entered the laboratory, handed the requested music to the researcher on the selected media, and was asked to sit down at a desk across from the listening booth. There, they found three forms in front of them on the desk: Background questionnaire (Appendix A.2), POMS mood assessment scale form, and an informed consent form. The participant was asked to fill out each of these while the researcher copied the music files onto the experimental computer in the listening booth. The researcher then readied the listening booth by connecting the first controller to the AVR-HID and then loading the first stimulus into the Max/MSP patch. The controller presentation was randomized using a Latin square configuration that offered six presentation orders. A participant was assigned a presentation order prior to the experiment in an incremental fashion. After Max/MSP was set up, a test was run by the researcher using a hidden part of the patch to determine if the hardware was working correctly.

Once the initial questionnaires were completed, the participant was directed into the listening booth and asked to sit down in the chair in front of the computer. The researcher handed them the first controller and the instructions for that particular controller (Appendix A.3, Appendix A.4, Appendix A.5). While reading the instructions, the participants were asked to hold the controller in hand and get a feel for its operation. (The participants were not shown the controllers before the experiment, so their introduction to the participant was based on the Latin square discussed earlier.) After the participant read the instructions, the researcher asked if they had any questions. Once proper use of the device was established, they were instructed on the following procedure: "To start the music, click the blank box on the screen. The music will stop automatically when the song is done. You may adjust the dial on the amplifier to whatever level you wish. Also, you may move the mouse pointer to the upper right corner of the screen for the screen to darken. After the song is over, please come out of the booth and fill out the next questionnaire on the desk. Would you prefer the lights on or off during the listening session?"

The researcher then adjusted the lights according to the participant's preference and left the booth. Once the participant came out of the booth, they filled out a Post-Experiment questionnaire (Appendix A.6). Then the cycle repeated from the connection of the next controller to the Post-experiment questionnaire for the last two stimuli. At the end, the researcher gave the participant the Post-Experiment questionnaire, paid them and thanked them for their cooperation. Finally, any questions that the participant had regarding the goals of the study or the theorized results were answered.

6. Results

6.1 Results

With the experiment examining many features including listener preferences, comfort levels, mood, and controller choice, this section is divided into several components to clarify the presentation of the results.

6.1.1 Music Education

Out of the participant pool, 83% had taken music instrument lessons, with 35% having private lessons, 10% having group lessons, and 55% having both. The mean age that lessons began was 7 years old (SD = 2.6). Not restricting to trained musicians, 79% of all participants play at least one instrument a mean time of 1.1 hours per day (SD = .49). The music players had several types of involvement in music making: 100% played alone, 47% play in a band, 95% play for fun, 10% play for school, and 42% play for profit. When asked if they compose original music, 74% of the music players answered in the affirmative.

6.1.2 Music Listening Habits

The pre-experiment questionnaire brought to light much information regarding the listening preferences of the participants. The mean listening frequency was found to be 3 hours per day (SD = 2.6), with the main motivation for listening listed as "emotional effect" by 58% of the participants, followed by 29% of the participants reporting "background music" as their main motivation. Only 8% of the subjects reported their primary motivation as "critical listening". The source of the music for 58% of the participants was playlists, with 71% of the playlists being hand-made versus an equal amount being either randomized or genre-generated. The rest of the participants (37%) listened to albums more often than playlists, with the exception of one participant who chose to select and listen to individual songs. The listening equipment preferred by the participants was loudspeakers for 67% of the participants. Headphone listeners were found to make up 29% of the participant pool, with one participant claiming an equal amount of use with both. Two-thirds of the participants preferred to listen to music alone as opposed to listening with friends. No one claimed to prefer listening to music with strangers. The only participants who listened to albums more than playlists or songs were those who preferred to listen to music alone. Participant responses of listening location had significant relationships with certain other listening habits. Participants who listen while traveling were more likely to be male ($p = .037$, Fisher's Exact Test) and were primarily playlist listeners ($p = .037$, Fisher's Exact Test). Participants listing concert listening as a primary location had a relationship to listening while travelling, with concert listening being selected each time travel listening was selected, and vice-versa. This listener background data will be analyzed alongside the mood data and the emotional response data below.

6.1.3 Mood Assessment

The Profile of Mood States (POMS) score is broken down into seven sections: Tension-Anxiety, Depression-Dejection, Anger-Hostility, Vigor-Activity, Fatigue-Inertia, Confusion-Bewilderment, and the Total Mood Disturbance score. These factors were narrowed down through six factor-analytic studies run by Loor, McNair, & Heuchert between the years of 1961 and 1965 and determined to be representative of six discrete mood factors that can be measured reliably with the derived test. The Total Mood Disturbance factor is determined as a single estimation of mood state. It is found by summing all factors except vigor-activity, then subtracting vigor-activity from the sum. This calculation was determined to be valid considering the positive correlations between the first five factors, and the negative correlation they all have with vigor-activity. The POMS results can be seen in Table 2.

Mood Factor (range)	**Minimum**	**Maximum**	**Mean**	**S.D.**
Tension-Anxiety (0-36)	1	9	4.4	2.5
Depression-Dejection (0-60)	0	28	5.1	6.4
Anger-Hostility (0-48)	0	9	2.3	2.6
Vigor-Activity (0-32)	7	22	14.5	4.3
Fatigue-Inertia (0-28)	3	17	6.5	3.6
Confusion-Bewilderment (0-28)	1	15	5.6	3.4
Total Mood Disturbance (-32-200)	-12	41	9.4	14.6

Table 2: POMS Results

6.1.4 Experimental Procedure Data

As discussed in the experimental procedure, the presentation of the controllers was randomized based on a Latin square in order to keep the presentation order from systematically affecting the controller preference data. Due to a mistake during the experiment, one of the presentation orders was duplicated, resulting in the frequencies of presentations as shown in Table 3. The controllers "Joystick", "Button", and "Slider" throughout the experiment were labeled as 1, 2, and 3, respectively.

Order of Presentation	1 2 3	1 3 2	2 1 3	2 3 1	3 1 2	3 2 1
Frequency	4	4	5	3	4	4

Table 3: Frequencies of Controller Presentation Order

Also, as discussed in section 3.4.3, the participants were asked if they wanted the lights in the booth turned off for the duration of the experiment. Of all 24 participants, 75% asked for the lights turned off. A chi-squared test comparing the frequencies of the listening motivation categories across the two light settings was highly significant ($X^2(3)=13.714$, $p = .003$), showing that 100% of the participants listing emotional effect as their primary motivation for music listening asked for the lights turned off, while 71% of the participants listing background music as primary motivation asked for the lights to remain on. Stimuli varied greatly in length from 46 s to 13 min 4 s (M = 4 min 48 s, SD = 2 min 17 s).

6.1.5 Joystick

The data obtained from ratings concerning the Joystick controller during the experiment is summarized in Table 4.

Joystick Questionnaire Responses					
	N	Minimum	Maximum	Mean	S.D.
Joystick Ease (1-7)	24	3	7	5.6	1.0
Joystick Comfort (1-7)	24	4	7	6.0	0.9
Joystick Accuracy (1-7)	24	2	6	4.8	1.1
Joystick Disturbance (1-7)	24	1	6	3.0	1.5
Recorded Joystick Data					
	N	Minimum	Maximum	Mean	S.D.
Joystick Range (0-1023)	23	242	1023	916	195
Joystick Minimum (0-1023)	23	0	513	93	155
Joystick Maximum (0-1023)	24	755	1023	1010	55
Joystick Mean (0-1023)	23	532	790	624	77
Std. Deviation	23	52	393	196	71
Number of Thrills	24	0	36	4.8	7.5

Table 4: Joystick Statistics

The top four rows are the scores that the Joystick controller received on the individual questionnaires after presentation of that specific controller. All the remaining rows are statistics regarding the data recorded from the controllers during the experiments. There is one data set excluded from the recorded data, with the exception of the maximum and number of thrills, due to an error that halted the recording after the first peak of the response, which was roughly 20 seconds into the piece. The maximum was retained because the fragmented data set still showed that the participant used a certain amount of the scale, and this data is relevant to the rest of the study.

Table 5 shows the correlations found in the Joystick questionnaire data:

		Joystick Ease	Joystick Comfort	Joystick Accuracy	Joystick Disturbance
Joystick Ease	Pearson Correlation	1	.33	**.69****	**-.54****
	p-value (2-tailed)	.	.114	**.000**	**.006**
	N	24	24	**24**	**24**
Joystick Comfort	Pearson Correlation	.33	1	**.41***	.00
	p-value (2-tailed)	.114	.	**.046**	.995
	N	24	24	**24**	24
Joystick Accuracy	Pearson Correlation	**.69****	**.41***	1	**-.51***
	p-value (2-tailed)	**.000**	**.046**	.	**.011**
	N	**24**	**24**	24	**24**
Joystick Disturbance	Pearson Correlation	**-.54****	.00	**-.51***	1
	p-value (2-tailed)	**.006**	.995	**.011**	.
	N	**24**	24	**24**	24

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Table 5: Joystick Correlations

Starting from the top, Joystick Ease was positively correlated with Joystick Accuracy and negatively correlated with Joystick Disturbance. Joystick Comfort had a positive correlation with Joystick Accuracy. Joystick Accuracy was negatively correlated with Joystick Disturbance.

6.1.6 Button

Table 6 shows the data gathered for the Button Controller:

Button Questionnaire Responses					
	N	Minimum	Maximum	Mean	S.D.
Button Ease (1-7)	24	3	7	5.2	1.2
Button Comfort (1-7)	24	4	7	5.9	1.1
Button Accuracy (1-7)	24	1	7	4.5	1.5
Button Disturbance (1-7)	24	1	7	3.2	1.5
Recorded Button Data					
	N	Minimum	Maximum	Mean	S.D.
Button Range	19	166	797	528	203
Button Minimum	19	226	857	495	203
Number of Thrills	22	0	27	5.4	7
Number of Presses	22	0	33	8.6	7

Table 6: Button Statistics

A few technical errors occurred during the use of the Button controller, which resulted in a full loss of two sessions, and a partial loss of two sessions as can be seen in the count on the bottom half of Table 6. Correlations within the Button dataset were found to be similar to those of the Joystick.

		Button Ease	Button Comfort	Button Accuracy	Button Disturbance	Subject's age
Button Ease	Pearson Correlation	1	**.45***	**.51***	**-.62****	-.30
	p-value (2-tailed)	.	**.027**	**.011**	**.001**	.153
	N	24	**24**	**24**	**24**	24
Button Comfort	Pearson Correlation	**.45***	1	.16	-.18	**-.49***
	p-value (2-tailed)	**.027**	.	.459	.410	**.015**
	N	**24**	24	24	24	**24**
Button Accuracy	Pearson Correlation	**.51***	.16	1	-.37	.10
	p-value (2-tailed)	**.011**	.459	.	.074	.658
	N	**24**	24	24	24	24
Button Disturbance	Pearson Correlation	**-.62****	-.18	-.37	1	.21
	p-value (2-tailed)	**.001**	.410	.074	.	.321
	N	**24**	24	24	24	24
Subject's age	Pearson Correlation	-.30	**-.49***	.10	.21	1
	p-value (2-tailed)	.153	**.015**	.658	.321	.
	N	24	**24**	24	24	24

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

Table 7: Button Correlations

As can be seen in Table 7, Button Ease correlated positively with Button Comfort, and Button Accuracy, and negatively with Button Disturbance. Button Comfort negatively correlated with participant age.

6.1.7 Slider

The Slider controller scores and statistics can be found in Table 8.

Slider Questionnaire Responses					
	N	Minimum	Maximum	Mean	S.D.
Slider Ease (1-7)	24	2	7	5.6	1.2
Slider Comfort (1-7)	24	2	7	5.2	1.4
Slider Accuracy (1-7)	24	2	6	4.7	1.2
Slider Disturbance (1-7)	24	1	6	3.5	1.5
Recorded Slider Data					
	N	Minimum	Maximum	Mean	S.D.
Slider Range (0-1023)	21	212	1023	684	205
Slider Minimum (0-1023)	21	0	588	228	185
Slider Maximum (0-1023)	21	624	1023	912	127
Slider Mean (0-1023)	21	409.40	883.40	602	118
Std. Deviation	21	69.32	300.80	165	53

Table 8: Slider Statistics

As can be seen in the second half of Table 8, 3 of the 24 Slider sessions were not recorded due to malfunction of the A/D converter. Some interesting connections arose during correlation analysis for the Slider. In Table 9, the correlations between the above variables are explored.

		Slider Ease	Slider Comfort	Slider Accuracy	Slider Disturbance	Slider Range	Slider Minimum	Slider Maximum
Slider Ease	Pearson Correlation	1	.25	.32	-.22	**.44***	-.03	**.66****
	p-value (2-tailed)	.	.247	.125	.295	**.048**	.892	**.001**
	N	24	24	24	24	**21**	21	**21**
Slider Comfort	Pearson Correlation	.25	1	.35	-.28	-.04	.14	.15
	p-value (2-tailed)	.247	.	.094	.177	.871	.540	.531
	N	24	24	24	24	21	21	21
Slider Accuracy	Pearson Correlation	.32	.35	1	**-.43***	-.22	.25	.01
	p-value (2-tailed)	.125	.094	.	**.037**	.336	.270	.963
	N	24	24	24	**24**	21	21	21
Slider Disturbance	Pearson Correlation	-.22	-.28	**-.43***	1	.48*	**-.59****	-.08
	p-value (2-tailed)	.295	.177	**.037**	.	.028	**.005**	.718
	N	24	24	**24**	24	21	**21**	21
Slider Range	Pearson Correlation	**.44***	-.04	-.22	**.48***	1	**-.79****	**.46***
	p-value (2-tailed)	**.048**	.871	.336	**.028**	.	**.000**	**.035**
	N	**21**	21	21	**21**	21	**21**	**21**
Slider Minimum	Pearson Correlation	-.03	.14	.25	**-.59****	**-.79****	1	.18
	p-value (2-tailed)	.892	.540	.270	**.005**	**.000**	.	.439
	N	21	21	21	**21**	**21**	21	21
Slider Maximum	Pearson Correlation	**.66****	.15	.01	-.08	**.46***	.18	1
	p-value (2-tailed)	**.001**	.531	.963	.718	**.035**	.439	.
	N	**21**	21	21	21	**21**	21	21

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

Table 9: Slider Correlations

Slider Ease positively correlated with Slider Range and Slider Maximum. Slider Accuracy showed negative correlations with Slider Disturbance. Slider Disturbance positively correlated with Slider Range and negatively correlated with Slider Minimum. Slider Range correlated positively with Slider Maximum and negatively with Slider Minimum. Also, independent-samples t-tests showed a strong relationship between the booth lights and Slider Comfort ($t(22)=3.7$, $p<.001$) and Slider Maximum ($t(19)=3.1$, $p<.006$): the means of these variables were significantly higher when the lights were on than when off.

6.1.8 Controller Comparison

Some interesting connections were discovered between the data from each controller. These findings are presented in Table 10.

		Joystick Comfort	Button Comfort	Slider Comfort
Joystick Comfort	Pearson Correlation	1	**.55****	**.50***
	p-value (2-tailed)	.	**.006**	**.013**
	N	24	**24**	**24**
Button Comfort	Pearson Correlation	**.55****	1	.26
	p-value (2-tailed)	**.006**	.	.224
	N	**24**	24	24
Slider Comfort	Pearson Correlation	**.50***	.26	1
	p-value (2-tailed)	**.013**	.224	.
	N	**24**	24	24

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Table 10: Controller Data Interaction Correlations

This table shows positive correlations between the rated comfort levels of all three controllers, with that between the Joystick and Button being slightly stronger than between the Joystick and Slider. Moving into the chills recorded by the participants on the Joystick and Button controllers, more connections became evident, as represented in Table 11.

		Number of Recorded Chills on Joystick	Number of Recorded Chills on Button	Button Range	Slider Disturbance	Slider Maximum
Number of Recorded Chills on Joystick	Pearson Correlation	1	.249	**.506***	**.523****	**-.504***
	p-value (2-tailed)	.	.263	**.027**	**.009**	**.020**
	N	24	22	**19**	**24**	**21**
Number of Recorded Chills on Button	Pearson Correlation	.249	1	.286	**.477***	-.336
	p-value (2-tailed)	.263	.	.235	**.025**	.160
	N	22	22	19	**22**	19
Button Range	Pearson Correlation	**.506***	.286	1	.157	-.154
	p-value (2-tailed)	**.027**	.235	.	.522	.542
	N	**19**	19	19	19	18
Slider Disturbance	Pearson Correlation	**.523****	**.477***	.157	1	-.084
	p-value (2-tailed)	**.009**	**.025**	.522	.	.718
	N	**24**	**22**	19	24	21
Slider Maximum	Pearson Correlation	**-.504***	-.336	-.154	-.084	1
	p-value (2-tailed)	**.020**	.160	.542	.718	.
	N	**21**	19	18	21	21

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

Table 11: Chill Response Correlations

As seen in Table 11, the chills recorded on the Joystick are positively correlated with Button Range and Slider Disturbance, while a negative correlation exists with Slider Maximum. Chills recorded on the Button only show one significant correlation, which was a positive relationship with Slider Disturbance.

6.1.9 Final Questionnaire

The last questionnaire, designed to allow the participants to make an informed selection about the easiest and the most accurate controller after using all three, did not establish any of the three controllers to be an indisputable favorite, but brought up some interesting issues. For the rating of easiest controller, each one received 8 votes, making the scores equal for this measurement. The selection of the most accurate controller, however, showed 46% of the participants selecting the Joystick, 42% the Slider, and 12% the Button. These differences show that the Joystick and Slider were preferred much more often than the Button with accuracy in mind.

The second part of the questionnaire turned out to be very significant, with connections stretching into many parts of the other data. These questions let the participants rate their comfort in the experiment, as well as how much the experimental setting interfered with their emotional response. Table 12 shows the responses to these questions.

	Minimum	Maximum	Mean	SD.
Interference	1	6	3.3	1.8
Comfort	2	7	5.8	1.4

Table 12: Experimental Setting Responses

As expected, the interference and comfort were negatively correlated ($r(23) = -.59$, $p<.002$). Important correlations were found when these two ratings were compared with the POMS results, as displayed in Table 13.

		Interference of Exp. Setting	Comfort of Experimental Setting	Depression-Dejection	Tension-Anxiety	Mood Disturbance
Interference of Exp. Setting	Pearson Correlation	1	**-.59****	**.41***	.30	.31
	p-value (2-tailed)	.	**.002**	**.044**	.154	.147
	N	24	**24**	**24**	24	24
Comfort of Experimental Setting	Pearson Correlation	**-.59****	1	**-.61****	**-.54****	**-.43***
	p-value (2-tailed)	**.002**	.	**.002**	**.006**	**.035**
	N	**24**	24	**24**	**24**	**24**
Depression-Dejection	Pearson Correlation	**.41***	**-.61****	1	.40	**.83****
	p-value (2-tailed)	**.044**	**.002**	.	.050	**.000**
	N	**24**	**24**	24	24	**24**
Tension-Anxiety	Pearson Correlation	.30	**-.54****	.40	1	**.48***
	p-value (2-tailed)	.154	**.006**	.050	.	**.019**
	N	24	**24**	24	24	**24**
Mood Disturbance	Pearson Correlation	.31	**-.43***	**.83****	**.48***	1
	p-value (2-tailed)	.147	**.035**	**.000**	**.019**	.
	N	24	**24**	**24**	**24**	24

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Table 13: Comfort Rating / POMS Correlations

Some of the strongest correlations were in the relation of comfort to the mood ratings. Tension-Anxiety, Depression-Dejection, and TMD were negatively correlated with the rated level of comfort in the experimental setting. Interference was also negatively correlated with comfort, and positively with Depression-Dejection.

There is a considerable amount of information available from the results regarding controller design, experimental design, and the effects on emotional response, all of which will be discussed in Chapter 7.

7. Discussion

7.1 Introduction

There were several goals for this study. The first was to implement novel modifications to established experimental methods and determine if they yielded positive, and useful, results. The second was to compare two new controllers used to record continuous emotional response to music with one previously used in several experiments. The third goal was to point out how young the field of evoked emotional response to music is, and therefore, to offer a number of factors yet to be studied. Due to the large number of results, the following discussion is divided into its relevant sections in order to clarify the methodological propositions and their outcomes in relation to previous hypotheses.

7.2 Questionnaires

The main questionnaire, found in Appendix A.2, was designed to include most of the standard questions (age, gender, music education) while inquiring thoroughly about each participant's habits involving listening and playing music. It employed many of the suggestions discussed in 3.6.1 in an attempt learn as much as possible about how they listen to music, where they listen to music, and to what kind of music they listen. The questionnaire was very straightforward, and no questions arose except for a few asking about how detailed they should describe the genres, e.g. "should I write 'rock' or 'post-punk/hardcore'."

As theorized in 3.2, the musical education of the participants had no effect on any aspect of the emotional responses recorded during the experiment, including the range of the controller use or the number of chills recorded by the participant. This supports both Bigand et al's (2003) and Gabrielsson's (2001) findings that emotional responses generally do not differ between musicians and nonmusicians. This does not mean that a classically trained musician who hears a very familiar piece of music will have the same emotional response as a hip-hop fan who has never heard it before. The finding speaks to the *ability* to have similarly powerful emotional responses. As each participant was asked to bring in three pieces of music that move them personally, this lack of correlation with any aspect of their musical education supports the theory that the amount of education does not dictate the possible power of an individual's emotional response to music. On a related topic, the prior prediction from section 3.3.1 that genres other than classical music can move people to strong emotional states was supported by the genres of the music brought by the participants. These stimuli, which were chosen by the participants to fit the criteria requiring familiarity and recurrent emotional force changes, were predominately nonclassical. Of all 24 participants, only 5 chose classical stimuli, supporting the earlier claim that classical music is not the only style of music that has the power to elicit emotional responses in listeners. Considering that 46% of the participants listed classical as one of their favorite genres (with no familiarity rating below 3 out of a possible 5), only half of them selected classical music for their "emotion-evoking" stimuli. While it is possible that this sample is unique, and would not represent the stimulus choices for other potential participants, it is useful to note that 83% had taken music lessons starting at a mean age of 7. Therefore most of the participants would potentially have more exposure to classical music than a sample made up of participants who had no musical training.

The listing of favorite genres on the questionnaire displayed both how many styles of music to which the participants listen and how meticulous some of the genre naming is. The questionnaires showed that 96% of the participants listed 3 or more genres as their favorite. Considering that most of the genres were rated at 3 or higher on the familiarity rating scale next to each genre line (only 5 participants rated any of their genre familiarity ratings below 2), it is likely that the participants actively listen to all of the genres they listed, so there is little chance that the participants felt urged by the demand characteristics of the questionnaire to fill all five available lines. With the participants having specific tastes spanning many genres, this finding, when combined with those of London (2002) and Gabrielsson (2002) regarding the importance of familiarity affecting emotional response, supports the need for researchers to record the participants' preferences in-depth, as described in 3.3.2. Familiarity on a temporal basis, as mentioned in 3.3.2, was found to be a factor in this experiment as well. A few of the participants communicated that they felt their recorded emotional response in the experiment was slightly less strong than the emotional response that they experienced while auditioning songs to bring to the experiment. This supports Grewe's (2005) finding that the number of chills recorded in subsequent listenings to the same song decreases with every presentation, until a large hiatus (on the order of days or weeks) was taken from the listening frequency. While this does not have a strong negative effect on the current experiment, mainly because the goal is to analyze the methods to record emotional responses not to analyze the actual emotional responses, it has the potential to lessen the possible magnitude of emotional responses in future studies. When stimulus selection is the duty of the researcher, not the participant, this artifact will be minimized, but it is best practice to question the participants about the last time they listened to the stimuli in order to account for any data anomalies later.

Considering music listening habits, some very interesting results came from the responses to the questionnaire. The motivation for listening to music showed a dichotomy between emotional effect (58%) and background music (29%). While these may not be complete polar opposites, they do represent active versus passive listening, therefore it is relevant for this and future studies that there were almost twice as many "emotionally-motivated" listeners as those who listened mainly for a sensory background to other tasks. Also, the relationship between motivation and choice of light settings in the listening booth was very revealing, with 100% of the participants who chose "emotional effect" as their primary motivation asking for the lights to be turned off. Ravaja and Kallinen (2004) found that music that triggered the startle response increased zygomatic facial muscle EMG levels (which are positively correlated with pleasure levels) when the participants had high scores on a Behavioral Activation System (BAS) evaluation, which can be generally regarded as extraversion. Considering that Grillon et al. (1997) found that darkened settings can increase the amplitude of startle response, a high-BAS person who listens to startling music in the dark would have a more emotional experience than one who listens with the lights on. This potential increase of the evoked emotion amplitude would explain why all participants who primarily listen to music for "emotional effect" asked for the lights to be turned off.

The majority's preference of loudspeakers over headphones and listening alone over listening with friends have no significant interrelation, yet have the potential to affect the experimental environment in future studies due to their possible experimental interactions discussed near the end of section 3.4: that headphones can enhance the mood by providing a separate experiential reality and that group conformity effects may distort self-report emotion ratings. Location of listening did not reveal too many surprising results, with the notable exception of those who selected listening while traveling as one of their primary locations. Their preference to listen alone may support Crane's (2005) theory that headphones act as a security blanket of sorts, that is, if the traveling is mainly done with the use of personal portable music devices. Travel listeners using primarily playlists is a relatively simple relationship mainly due to the convenience of personal portable music devices, which are very often playlist driven, and the lack of personal album collections while traveling. The fact that no female participants listed travel listening as one of their locations is most likely due to the unbalanced gender makeup of the participant pool, so further research would be needed to determine if males truly listen while traveling more often than females. Though, it should be noted that the question was not simply, "where do you listen to music?" but "where do you listen to music for the above primary motivation?" With that in mind, it is possible that men apply their primary listening motivation to traveling more often than women, which would be supported by findings in Vignoli (2004) that found that men had significantly larger digital music collections than women, thus giving men the advantage to bringing the appropriate music with them during travel.

7.3 Experimental Environment

As mentioned in the experimental procedure, the soundproof booth was set up in order to make it as comfortable as possible to ease the participant into a comfort level conducive to emotional responses. To aid this pursuit, the participants were allowed to choose the setting of the lights (on/off), the setting of the computer screen (on/off), and the volume. Even with these variables, the participants still did not feel completely comfortable. The three questions at the end of the final questionnaire (Appendix A.7) attempted to record the level of interference and comfort, as well as give the participant a chance to offer their suggestions to improve comfort in a free response question. As can be seen in Table 13, the strong negative correlation between Comfort of Experimental Setting and Depression-Dejection, Tension-Anxiety, and TMD conveys that the mood of the participant, more importantly the participant's levels of depression and tension, will make the experimental setting feel less comfortable. This in turn may affect the magnitude of the emotional response, which supports the use of mood assessment as a methodological practice before (and possibly after) an emotivist experiment is run.

In the final question, many participants had comments about the setting and how to make it more comfortable. Lighting concerns were mentioned by six of the participants, most of which referred to how much the lights on the monitor system dB readout bothered them, as well as the power lights on the monitor. These subjects did have the lights in the booth turned off, so that amplified the intensity of the glowing blue lights to a point that bothered many. Three other participants wrote that the computer screen, keyboard, and mouse were distracting because they did not need to use them during the experiment (with the exception of starting playback). Others talked of unfamiliarity and experimental pressure, explaining that "it felt too impersonal" and an experiment is "discomfiting at any place." In similar responses, participants wrote that it takes time to get used to the setting and that there should be "practice songs" to get used to the environment. One participant said that the "last evaluation was easier, perhaps because it takes time to get used to the setting." These many concerns support the need to take these experiments out of the clinical setting if absolutely genuine emotional responses are desired. At this point, the setting's direct effect on the magnitude of the emotional response is not known, yet two-thirds of the participants complaining about similar problems with the setting amplifies the severity of the potential problem, and necessitates further research regarding its effects.

7.4 Controllers

As discussed earlier, one of the main goals of this thesis is to explore new controller designs in comparison to an established one through the analysis of user ratings and use patterns. The following sections will discuss the controllers separately then culminate in a cross-comparison.

7.4.1 Joystick

The instructions given to the participants for the Joystick were simple: move the stick up when you feel an upward movement in your emotional force or down when you feel a downward emotional force, and hold it there until the change in force ends. The participants were never informed whether to move it partially or fully to the north and south poles. This ambiguity was deliberate in order to study the natural instinct the participants would have with the device. As expected, the majority of the participants did not use the Joystick in a binary manner. Of all 24 sessions, only three participants used three positions of the Joystick: up, down, and center. All of the rest used it just as any other linear controller; they still pressed up when the emotional change went up, and down when the change went down, but the magnitude of the press varied along with the intensity of the change. As one participant wrote on the questionnaire (Appendix A.6), "the push and pull of the joystick seemed parallel to my emotional response." Another very interesting use pattern was the predominance of upward presses versus downward. In every session, there was at least double the number of upward peaks than downward peaks. The Joystick Mean data in Table 4 support this finding with the minimum Mean being greater than the halfway point (512) between the two poles. This is likely due to a combination of two factors. First, the stimuli were chosen by the participants to have a strong emotional response, therefore many more upward changes than downward would be expected. Second, it may be easier and more natural to recognize an upward change in emotional force than a downward change, so more of them were recorded. The Joystick Maximum and Joystick Minimum values support this as well, considering that only 2 of 24 participants did not hit the maximum, while 9 of 24 did not hit the minimum.

As for user ratings, as the participants found the Joystick easier to use, they also rated it as more accurate and less disturbing of the emotional experience. Rated accuracy also increased as disturbance ratings decreased and as comfort ratings increased. These findings, while very straightforward, show that ease of use and comfort can have serious effects on the participant's confidence in controller accuracy, as well as the frequency of emotional disturbances. They also help to validate the four-dimensional rating system within this controller evaluation context. The four questions, regarding comfort, ease, perceived accuracy, and emotional disturbance, were designed to be interrelated. It was predicted that ease, comfort, and accuracy would correlate positively with each other, while disturbance would correlate negatively with all three. Therefore, the Joystick's and

Button's ratings showed that the rating system itself has merit for future use due to the moderately high level of correlation between the dimensions.

User feedback for the Joystick had two main themes: force feedback and "thrill" button location. Different modes of force feedback were recommended, including "time-dependent" and "location-dependent" force feedback. In the first case, the stick would become increasingly harder to hold in a direction as the participant held it there for a longer period of time. The second would be a more standard type of force feedback in which the participant would have to press the stick increasingly harder to move it to the extremes of the scale. The "thrill" button location suggestions were simply that having one hand control the stick and the other press the "thrill" button would be more convenient and eliminate the time required to "get used to coupled fingers."

7.4.2 Button

The Button was designed with the instructions of the Joystick in mind, but would only allow the recording of upward changes in emotional force. It was built with a force sensitive resistor (FSR) underneath the button, yet the instructions (Appendix A.4) were similar to the Joystick in ambiguity regarding the amount of pressure applied to the Button. This was to determine whether a participant would naturally press it harder when the emotional force change was more intense, just as Clynes' (1977) research would suggest. Unfortunately, the FSR was not sensitive enough to pick up the lighter touches, so the stronger presses were recorded correctly with some of the lighter ones being missed. Yet, as stated before, the response itself is not important in this study; to the participant, all was recorded on every press, and therefore, they could rate accurately, assuming that the equipment worked flawlessly. Visual analysis of the data did show that multiple pressures on the button were employed during most of the sessions, thus further research is required to determine if the different pressures are musically related or just coincidental.

User ratings show similar correlation patterns to those of the Joystick for the four-dimensional rating scale: Button Ease increases with Button Comfort and Button Accuracy, while Button Disturbance decreases. Interestingly enough, the age of the participant decreases as Button Comfort increases, which may be related to the construction of the controller, which was built from video game controller parts. This would mean that the younger participants are more comfortable with the Button due to video game playing habits that may be absent in the older population.

User comments about the Button centered around two concerns: force-sensing ability and "thrill" button location. As no participant knew that the button was sensitive to force, almost every participant wrote that "linear scales are better than binary" or "the button does not capture nuance," recalling the way that the Joystick was instinctively used by most participants. The "thrill" button was used more often on the Button than the Joystick (18 participants pressed the button at least once on the Button versus 15 on the Joystick), yet similar responses suggested, "two handed control would help." Some participants also "got confused between buttons." This is understandable because the modalities are so similar that confusion may present itself in a different way than the Joystick, which offers two modalities.

7.4.3 Slider

Slider design and instructions were identical to the study in which it was first implemented (McAdams, et al., 2004), so that its established methodology would not be altered and the data could potentially be compared across studies. This meant an emotional recording paradigm shift for the participants, with the slider employing an absolute scale. Instead of recording the locations of the changes in emotional force, the participants needed to constantly monitor their level of emotional force, and use the slider to represent its location from weak to strong. The Latin Square order of controller presentation attempted to remove any effect caused by this change in instructions.

When analyzing the use patterns of the Slider, Table 8 shows that the full range of the Slider was not used by most of the participants. The mean Slider Range is only slightly more than half of the scale, yet the mean Slider Maximum is considerably high, so the range used was in the upper end of the scale. This is most likely accounted for by the fact that the stimuli were especially emotionally arousing for each participant, thus not requiring much use of the "weak" end of the scale. At first glance, this may not seem universally true, considering that 6 of 21 participants had a Slider Maximum around three-quarters of the full scale. An assumption could be made that personally-chosen emotion-evoking music would necessitate the full use of the strong side of the scale, yet, these anomalies may be explained by the mood assessment. A slight negative correlation ($r(20)=.48$, $p<0.03$) was found between Slider Maximum and Fatigue-Inertia, which could account for weaker emotional force ratings.

User ratings did not resemble the other two controllers in the four-dimensional rating scale. The only correlation found in these was a somewhat weak negative relationship between Slider Accuracy and Slider Disturbance. Interestingly, Slider Ease was discovered to increase along with Slider Maximum and Slider Range, showing that when participants used the higher end of the scale, they rated the Slider to be easier to use. This may be related to familiarity with the personal emotional response signature of the stimuli. For example, if a participant is very familiar with the emotional peaks and valleys that they experience with a certain stimulus, then they may find it easier to use the slider due to the lack of scale location re-evaluation that would be needed if the song was not as familiar, as discussed in 4.7.3. A second cause may simply be that if the participant found the interface of the Slider to be easy to use, they would be more apt to use the full range to record their evoked emotions. Another significant relationship with these Slider attributes is that the setting of the lights in the booth affected both the Slider Comfort rating and the Slider Maximum. So, when the lights were on, the Slider felt more comfortable to use and the participants used more of the Slider. This is not very surprising, when the darkness could easily obscure the visual component necessary to use the Slider to its full potential. The participants in the dark may therefore have been reticent to move the Slider too far to its positive side to keep from hitting the maximum. This effect then accounts for lower Slider Ease ratings, because the participants felt it difficult to use in the dark, even though the dark is the most comfortable setting for their optimum flow of emotions.

User feedback for the Slider brought many new issues to light, as well as highlighted issues found in the pilot experiment. As before, participants wrote that they "had to open eyes to look at limits," that it was "difficult to use with a max and min," and that "every time I looked at it, I lost sensation." Many asked for ergonomics to be considered, claiming that it was awkward to hold. Some suggested that an interface that a participant could "squeeze or apply pressure" to would "reduce visual distraction."

Another set of relevant comments concerned the lack of a "thrill" button. One participant "missed being able to punctuate the emotional effect with additional input [i.e. "thrill" button]" while another stated, "I wish this device could record chills like the others." While every response about the lack of a "thrill" button was written by participants who used either the Button or the Joystick before using the Slider, it still is important to note that these participants felt that a "thrill" button would make the recording of emotional response easier and less distracting. This could be validated in future studies by offering each participant the choice of using the button, and recording how many take advantage of it.

7.4.4 Cross-Controller Comparison

When comparing the data between the controllers, a few interesting correlations are found between the ratings. The positive relationship found between Joystick Comfort and the comfort ratings of the other two controllers is somewhat surprising, though the stronger correlation with Button Comfort than Slider Comfort is not. Both the Joystick and Button controllers were built from the same hardware, so they should be rated with similar levels of comfort. The correlation between Joystick and Slider comfort, on the other hand, is very possibly a coincidence, as the methods of use for the two devices, e.g. hand position, finger movement, etc., are so radically different that the relationship would seem be due to a chance occurrence. The number of "thrills" recorded on the Button and Joystick were related to a number of other variables. Most importantly, as more "thrills" were recorded on either device, the rating of Slider Disturbance increased. This relates directly to the comments about the Slider's lack of "thrill" recording capabilities, as well as the need for uninterrupted focus on the music required for "thrill" production. For a participant that experiences many "thrills," the attention needed for Slider operation greatly interrupted their emotional experience and did not allow for any recording of any "thrills," thus resulting in higher Slider Disturbance ratings.

One experimental factor that could affect these ratings, and would most likely be present in any study that used pre-selected musical stimuli, is that of participants lacking familiarity with the stimuli. The design of this study was focused on maximizing the emotional response in the participants in order to examine their use of the controllers during high levels of emotional evocation. Therefore, having participants select the stimuli was the best choice to ensure emotional responses in all participants. Once a participant is asked to rate their emotional response to music they have never heard before, the controller use may differ, most notably in the Slider, due to the constraints of the absolute scale. In the case of this study, the participants knew the music and their typical response well, and therefore knew when to expect peaks and valleys in their response. Yet, even having this familiarity, many participants were interrupted by the Slider due to its physical limits. If they had no expectations about the form of their response to the stimulus, the disturbance has great potential to increase, as they would be forced to keep the Slider from the top position, as discussed in 4.7.3.

Stepping back, it can be seen that the Joystick controller met the criteria established in 4.7, which had been gleaned from in-depth analysis of all previous continuous response controllers. It allowed for transparent encoding of emotion by being "gesturally relevant". This is seen not only by the recorded data, which showed varied, yet full, use of the range as well as the frequent use of the "Thrill" button, but also through the comments supplied by participants. It was never criticized in comments for any type of emotional interruption, and was often praised for its comfort and ergonomics. It also did not require any visual component to operate, thus making it ideal for the darkened situation that 75% of the participants preferred. It is not, however, a perfect controller. Force feedback is a valid option for future controllers, and one that was requested often by the participants. Wireless transmission is also a possible improvement that would facilitate its use outside of the laboratory setting.

Another very important result was the desire that the participants had for a variable transducer. The most prominent comments spanning the Joystick and the Button revolved around the preference for the pressure-sensitive Joystick and the dislike of the seemingly "binary" Button. These comments were supported by Joystick graphs that displayed the participants pressing the Joystick further in a direction if their emotional response increased with more intensity. This use pattern was almost universal, despite never explicitly instructing the participants to use the intermediate positions in the Joystick range. Such a lack of explicit instructions would not be acceptable within any other emotivist study, but was required here to examine the instinctual use of the device.

This desire for pressure-sensitive rating scales may also account for the 88% of participants who did not find the Button to be the most accurate. They believed that it was a simple on/off button, and commented several times that such a device could not have ever recorded their complex emotional response accurately.

The enjoyment of an absolute scale to rate the intensity of each individual change in emotional force did not, however, result in favorable reactions to the absolute scale for rating total level of emotional force, as on the Slider. For example, one participant, after choosing the Joystick as the easiest and most accurate, wrote, "[The Joystick] didn't have a maximum/minimum as did the Slider, and it had a range, unlike the Button." This clearly states a desire for an unbounded scale for the recording of changes and an absolute scale that records the intensity of each change.

There also was a subset of participants who expressed much enthusiasm for the method of recording only increases in emotional force. Their desire for this stems from self-professed lack of ability to correctly identify when emotional force decreases. This was hypothesized prior to the controller design, and was the impetus for the Button controller. This remark, however, was always expressed with the need for the scale to be variable, just like the Joystick. One participant even combined the Button and Joystick designs in a suggestion to confine the Joystick to its positive y-axis and only rate upward changes with it.

8. Conclusion

This thesis had two main goals from the outset: 1) Propose an experimental methodology that takes into consideration the cognitivist or emotivist goal of the study and advise about the many caveats such studies warrant. 2) Design, build, and test two new controllers and compare their usability against an already established controller within an emotivist experiment, while implementing a subset of the guidelines proposed in the aforementioned methodology. These goals have been accomplished and much information has been found regarding all the aspects of experimental design and continuous recording of evoked emotions.

The experiment was exploratory in nature, with the intention of surveying the participants about everything related to their music-based habits, even though there would not be (within the scope of this thesis) an analysis of each participants' emotional response to their individual stimuli. Its importance was derived from the significantly varied responses that arose in almost all areas, including motivation for listening, the environment in which the participants feel comfortable listening, and the large number of genres to which they listen. Based on the research cited in the methodology proposal, these factors can have a large impact on personal emotional responses, and knowing their variance will drive researchers to study each factor individually, for which this methodology allows.

In the realm of controllers, a great amount of valuable information was gathered that will allow the refinement of future designs. There was a universal desire to rate the intensity of the emotional force change on an absolute scale, while rating the length of the change on an unbounded scale. Ergonomics were commented on frequently, concerning the high level of comfort of the Joystick and Button, along with the need for a change in the Slider's somewhat large casing.

Future research in this area is daunting, but very promising. As long as researchers clearly identify their studies as emotivist or cognitivist, and utilize strict, consistent research methodologies, comparison of results will facilitate the progress needed in this field. Controllers can be designed with detailed information concerning usability, and rating systems such as the four-dimensional rating system (ease, comfort, accuracy and emotional disturbance) used in this experiment will allow direct comparison between studies. Research into force feedback integration for controllers must also be at the forefront, considering its prominence in the comments within this study.

Finally, one of the most important developments in this field is the inclusion of physiological measurements, such as galvanic skin response for measuring arousal and electromyogram of facial muscles to record emotional valence. The combination of continuous response controllers and physiological measurement will provide invaluable information about not only how the mind responds to music emotionally, but also how our body physically reacts to it, which will let researchers take another step towards a unified theory of emotional response to music.

Appendix A: Questionnaires and Instructions

A.1 Pilot Experiment Questionnaire

What is the number on your Emotional Force Slider Box? _____

What was your overall emotional valence response to each piece? (circle one)

1. (positive / negative)
2. (positive / negative)
3. (positive / negative)
4. (positive / negative)
5. (positive / negative)

How familiar are you to the styles of music presented today? (circle one)

(very familiar / moderately familiar / unfamiliar)

Do you listen to this style of music outside of the experimental environment?

Would judging emotional valence be easier or harder than judging emotional force in the context of this experiment?

How easy was it to control the slider in such a way that your emotional force was recorded transparently?

A.2 Background Questionnaire

Pre-Experiment Questionnaire

Gender: (Male / Female)

Age:____

Handedness: (Right / Left)

<u>Musical Training:</u>

Have you taken music lessons? (Yes / No)

If yes:

At what age did you begin lessons? ____

Private or group lessons? ________

Which instruments? How many years?

____________________ __________

____________________ __________

____________________ __________

____________________ __________

Do you currently play an instrument? ____________ How often? (hours per day) _______

What style(s) of music do you play currently? ____________

Do you play your instrument(s): (circle all that apply)

Alone With a band/group For fun For school For profit

Do you write original music? (yes / no)

<u>Musical Experience:</u>

How often do you listen to music? (hours per day) _________

What is your primary motivation for listening to music? (circle one)

Background music Critical listening Emotional effect Other:______________

Where do you usually listen to music for the above motivation? (circle all that apply)

Home Work Travel Concerts

Do you listen more often to playlists, albums, or individual songs? _____________

If playlists: (Hand-made / Genre-generated / Purely randomized)?

Do you listen more often on: (Headphones / Speakers)?

Do you prefer to listen to music: (Alone / With friends / With strangers)?

What are your favorite styles of music?	How well do you know the style?				
______________________________	1	2	3	4	5
______________________________	1	2	3	4	5
______________________________	1	2	3	4	5
______________________________	1	2	3	4	5
______________________________	1	2	3	4	5

A.3 Joystick Instructions

Instructions for Joystick Controller

You will hear a piece of music for which you have previously experienced emotional responses. During the piece, you will rate your experience of emotional force as the music progresses. It is important to remember that you should only rate the changes in your own emotional force, not in any emotions that you may 'hear' or 'perceive' in the music. Also it is the force of this emotional reaction that you will rate and not its positive or negative nature or character. For example, you can rate a feeling of joy or exaltation as having the same force as a feeling of anguish or sadness.

You will make these ratings on the given controller. Hold the controller in whichever hand feels most comfortable. When you feel your emotional force getting stronger, press the joystick upwards. When you feel your emotional force getting weaker, press the joystick downwards. Keep the joystick pressed for the duration of the change in emotional force.

There is also a trigger that you will operate with your index finger during the experiment. Press and hold the trigger when you feel any kind of physiological reaction from the music, e.g. chills up/down your spine, lump in your throat, tears, etc.

There are no right or wrong answers. What matters to us is how <u>you</u> react emotionally to the music. **If you have any questions, please do not hesitate to ask them before the end of the explanation session.**

A.4 Button Instructions

Instructions for Button Controller

You will hear a piece of music for which you have previously experienced emotional responses. During the piece, you will rate your experience of emotional force as the music progresses. It is important to remember that you should only rate the changes in your own emotional force, not in any emotions that you may 'hear' or 'perceive' in the music. Also it is the force of this emotional reaction that you will rate and not its positive or negative nature or character. For example, you can rate a feeling of joy or exaltation as having the same force as a feeling of anguish or sadness.

You will make these ratings on the given controller. Hold the controller in whichever hand feels most comfortable. When you feel your emotional force getting stronger, press the button located under your thumb. Keep the button pressed for the duration of the change in emotional force. You are only rating when your emotional force gets stronger.

There is also a trigger that you will operate with your index finger during the experiment. Press and hold the trigger when you feel any kind of physiological reaction from the music, e.g. chills up/down your spine, lump in your throat, tears, etc.

There are no right or wrong answers. What matters to us is how you react emotionally to the music. **If you have any questions, please do not hesitate to ask them before the end of the explanation session.**

A.5 Slider Instructions

Instructions for Slider Controller

You will hear a piece of music for which you have previously experienced emotional responses. During the piece, you will rate your experience of emotional force as the music progresses. It is important to remember that you should only rate the changes in your own emotional force, not in any emotions that you may ‘hear’ or ‘perceive’ in the music. Also it is the force of this emotional reaction that you will rate and not its positive or negative nature or character. For example, you can rate a feeling of joy or exaltation as having the same force as a feeling of anguish or sadness.

You will make these ratings on the given controller. You have a black box with a slider that can be moved horizontally from "Weak" on the left to "Strong" on the right. The position of the slider should reflect at each moment the force of your emotional response to the piece as you are listening. To translate continuously this force, you will need to be constantly monitoring the force of your emotional response to what you are hearing in order to keep the slider at the position corresponding to these feelings.

There are no right or wrong answers. What matters to us is how you react emotionally to the music. **If you have any questions, please do not hesitate to ask them before the end of the explanation session.**

A.6 Post-Controller Questionnaire

Post-Experiment Questionnaire

How easy was it to rate the changes in your emotional force on the provided device?

Difficult	1	2	3	4	5	6	7	Easy

How comfortable was it to use the device?

Uncomfortable	1	2	3	4	5	6	7	Comfortable

How accurately do you feel the device recorded your changes in emotional force?

Inaccurately	1	2	3	4	5	6	7	Accurately

How frequently did the use of the device interrupt your emotional experience?

Infrequently	1	2	3	4	5	6	7	Frequently

Are there any ways that the device could be changed to allow you to record your changes in emotional force more easily and in a way that would distract you less from listening?

____

____

____

A.7 Post-Experiment Questionnaire

Post-Experiment Questionnaire

Of the three controllers presented:

Which did you find easiest to use?

1 2 3

Which do you feel recorded your changes in emotional force most accurately?

1 2 3

Any comments on your above choices?

____

____

____

How much did the setting of the experiment interfere with your emotional experience?

Not at all 1 2 3 4 5 6 7 Greatly

How comfortable were you in the experimental setting?

Uncomfortable 1 2 3 4 5 6 7 Comfortable

Is there anything you would change about the setting to make it more comfortable?

____

____

____

Bibliography

Asch, S. E. (1951). Effects of group pressure upon the modification and distortion of judgments. In H. Guetzkow (Ed.), *Groups, leadership and men; research in human relations* (pp. 177-190). Oxford, England: Carnegie Press.

Asch, S. E. (1956). Studies of independence and conformity: I. A minority of one against a unanimous majority. *Psychological Monographs, 70*(9), 1-70.

Bigand E. (2003). More about the musical expertise of musically untrained listeners. *Ann. NY Acad. Sci.* 999: 304–12.

Blood, A. J., & Zatorre, R. (2001). Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion. *Proceedings of the National Academy of Sciences, USA*, *98*, 11818-11823.

Brittin, R. V. (1991). The effect of overtly categorizing music on preference for popular music styles. *Journal of Research in Music Education*, *29*, 47-56.

Capperella, D. A. (1989). Reliability of the Continuous Response Digital Interface for data collection in a study of auditory perception. *Southeastern Journal of Music Education, 1*, 19-32.

Clynes, M. (1977). *Sentics: The touch of emotions.* New York: Doubleday.

Crane, R. (2005). Social distance and loneliness as they relate to headphones used with portable audio technology. M.A. thesis, Humboldt State University, California.

Davis, W. B., & Thaut, M. H. (1989). The influence of preferred relaxing music on measures of state anxiety, relaxation, and physiological responses. *Journal of Music Therapy, 26*, 168–187.

DeVries, B. (1991). Assessment of the affective response to music with Clynes' sentograph. *Psychology of Music, 19*, 46-64.

Dibben, N. (2004). The role of peripheral feedback in emotional experience with music. *Music Perception, 22*(1), 79-115.

Gabrielsson, A. (2001). Emotions in strong experiences with music. In: P.N. Juslin, & J.A. Sloboda (Eds), *Music and Emotion: Theory and Research*, Oxford: Oxford University Press: 431–452.

Gabrielsson, A. (2002). Emotion perceived and emotion felt: Same or different? *Musicae Scientiae. Spec Issue, 2001-2002,* 123-147.

Gerra, G., Zaimovic, A., Franchini, D., Palladino, M., Guicastro, G., Reali, N., Maestri, D., Caccavari, R., Delsignore, R., & Brambilla, F. (1998). Neuroendocrine responses of healthy volunteers to "techno-music": relationships with personality traits and emotional state. *Int J Psychophysiol 28*, 99-111.

Goldstein, A. (1980). Thrills in response to music and other stimuli. *Physiological Psychology, 8,* 126-129.

Gregory, D. (1989). Using computers to measure continuous music responses. *Psychomusicology, 8*, 127-134.

Grewe, O., Nagel, F., Kopiez, R., & Altenmuller, E. (2005). How does music arouse "chills"? Investigating strong emotions, combining psychological, physiological, and psychoacoustical methods. *Ann NY Acad Sci, 1060,* 446-449.

Grillon, C., Pellowski, M., Merikangas, K.R., Davis, M. (1997). Darkness facilitates the acoustic startle reflex in humans. *Biol. Psychiatry, 42*, 461-471.

Filipic, S. & Bigand, E. (2004). The time-course of emotion and cognition while listening to music. In 8th International Conference on Music Perception and Cognition, Northwestern University.

Finaas, L. (2001). Presenting music live, audio-visually or aurally - does it affect listeners' experiences differently? *British Journal of Music Education 18*(1), 55-78.

Hevner, K. (1936). Experimental studies of the elements of expression in music. *The American Journal of Psychology. 48*(2), 246-268.

Juslin, P. N., & Zentner, M. R. (2002). Current trends in the study of music and emotion: Overture. *Musicae Scientiae. Spec Issue, 2001-2002*, 3-21.

Kivy, P. (1990). Music alone: Philosophical reflections on the purely musical experience. Ithaca: Cornell University Press.

Krumhansl, C., (1997). An exploratory study of musical emotions and psychophysiology. *Canadian Journal of Experimental Psychology, 51*, 336-353.

London, J. (2002). Some theories of emotion in music and their implications for research in music psychology. *Musicae Scientiae. Spec Issue, 2001-2002*, 23-36.

Lowis, M. J. (1998). Music and peak experiences: An empirical study. *Mankind Quarterly, 39,* 203-224.

Lychner, J. (1998). An empirical study concerning terminology relating to aesthetic response to music. *Journal of Research in Music Education, 46*, 303–320.

Madsen, C. K. (1997). Emotional response to music as measured by the two-dimensional CRDI. *Journal of Music Therapy, 34,* 187-199.

Madsen, C. K. (1998). Emotion versus tension in Haydn's symphony no. 104 as measured by the two-dimensional continuous response digital interface. *Journal of Research in Music Education, 46*(4), 546-554.

Madsen, C. K. & Fredrickson, W. E. (1993). The experience of musical tension: A replication of Nielsen's research using the Continuous Response Digital Interface. *Journal of Music Therapy, 30*(1): 46-63.

Madsen, C. K., & Geringer, J. M. (1990). Differential patterns of music listening: Focus of attention of musicians versus nonmusicians. *The Bulletin of the Council for Research in Music Education, 105*, 45-57.

Madsen, C. K., Brittin, R. V., & Capperella-Sheldon, D. A. (1993). An empirical method for measuring the aesthetic experience to music. *Journal of Research in Music Education, 41*(1), 57–69.

Madsen, C. K., Byrnes, S. R., Capperella-Sheldon, D. A., & Brittin, R. V. (1993). Aesthetic response to music: Musicians versus nonmusicians. *Journal of Music Therapy, 30,* 174-191.

Madsen, C. K., Capperella-Sheldon, D. A., & Johnson, C. M. (1991). Use of the continuous response digital interface (CRDI) in evaluating music responses of special populations. *Journal of the International Association of Music for the Handicapped* (formerly *MEH Bulletin*), *6*(2), 3-15.

Matsuura, M. (1998). Effects of background music on differential digital photoplethysmograms and mood states. *Japanese Journal of Physiological Psychology and Psychophysiology, 16*, 13-23.

McAdams, S., Vines, B. W., Vieillard, S., Smith, B. K., & Reynolds, R. (2004). Influences of large-scale form on continuous ratings in response to a contemporary piece in a live concert setting. *Music Perception, 22,* 297–350.

McLeod, K. (2001). Genres, subgenres, sub-subgenres and more: Musical and social differentiation within electronic/dance music communities. *Journal of Popular Music Studies, 13*(1), 59-75.

Meyer, L.B. (1956). *Emotion and meaning in music*. Chicago: University of Chicago Press.

Nagel, F., Grewe, O., Kopiez, R. & Altenmüller, E. (2005). *The relationship of psycho-physiological responses and self-reported emotions while listening to music.* Proceedings of the 30th Göttingen Neuro Biology Conference, University of

Göttingen, Germany, 17-20 February [CD-ROM].

Nielsen, F. V. (1983). *Oplevelse af musikalsk spœnding [The experience of musical tension]. Copenhagen: Akademisk Forlag.*

Orne, M. T. (1959). *The demand characteristics of an experimental design and their implications.* Paper read at American Psychological Association, Cincinnati.

Orne, M.T. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist, 17*, 776-783.

Panksepp, J. (1995). The emotional source of "chills" induced by music. *Music Perception, 13,* 171-207.

Panksepp, J. and Bernatzky, G. (2002) Emotional sounds and the brain: The neuro-affective foundations of musical appreciation. *Behav. Processes, 60,* 133–155

Ravaja, N. & Kallinen, K. (2004). Emotional effects of startling background music during reading news reports: The moderating influence of dispositional BIS and BAS sensitivities. *Scandinavian Journal of Psychology, 45,* 231-238.

Rentfrow, P. J., & Gosling, S. D. (2003). The do re mi s of everyday life: The structure and personality correlates of music preferences. *Journal of Personality and Social Psychology, 84,* 1236-1256.

Russell. J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology, 39.* 1161-1178.

Scherer, K. R., & Zentner, M. R. (2001). Emotional effects of music: Production rules. In P. N. Juslin & J. A. Sloboda (Eds.), *Music and emotion: Theory and research* (pp. 361–392). Oxford, England: Oxford University Press.

Schubert, E. (1996). Continuous response to music using a two dimensional emotion space. *Proceedings of the 4th International Conference of Music Perception and Cognition.* Montreal, Canada, 263-268.

Schubert, E. (1999) Measurement and time series analysis of emotion in music. PhD thesis, University of New South Wales.

Sloboda, J. A. (1991). Music structure and emotional response: Some empirical findings. *Psychology of Music, 19,* 110-120.

Sloboda, J. A. (2002). The "sound of music" *versus* the "essence of music": Dilemmas for music-emotion researchers (Commentary). *Musicae Scientiae. Spec Issue, 2001-2002*, 237-255.

Smith, J. L., & Noon, J. (1998). Objective measurement of mood change induced by contemporary music. *Journal of Psychiatric and Mental Health Nursing, 5*, 403–408.

Tyler, P. (1996). Developing a two-dimensional continuous response space for emotions perceived in music. Unpublished doctoral dissertation, Florida State University, Tallahassee.

Viellard, S., Bigand, E., Madurell, F., & Marozeau, J. (2003). The temporal processing of musical emotion in a free categorization task. *Proceedings of the 5th Triennial ESCOM Conference, Germany*, 234-237.

Vignoli, F. (2004). Digital Music Interaction Concepts: A user study. In Proc. 5th International Conference on Music Information Retrieval.

Winkielman, P., Knauper, B., & Schwarz, N. (1998). Looking back in anger: Reference periods change the interpretation of emotion frequency questions. *Journal of Personality and Social Psychology, 75,* 719-728.

Zentner, M. R., Meylan, S., & Scherer, K. R. (2000). Exploring 'musical emotions' across five genres of music. Paper presented at the Sixth International Conference of the Society for Music Perception and Cognition (ICMPC), 5-10 August 2000, Keele, UK.