AN EVALUATION OF THE EFFICACY OF DIGITAL REAL-TIME NOISE CONTROL TECHNIQUES IN EVOKING THE MUSICAL EFFECT

A.T. Warneke

2012

AN EVALUATION OF THE EFFICACY OF DIGITAL REAL-TIME NOISE CONTROL TECHNIQUES IN EVOKING THE MUSICAL EFFECT

Andrew Warneke

Submitted in partial fulfilment of the requirements for the degree of Master's in Music (Music Technology) to be awarded at the Nelson Mandela Metropolitan University

December 2012

Supervisor: Dr R. Bower

Co-Supervisor: Mr M. Brand

DECLARATION BY CANDIDATE

DECLARATION

I, *Andrew Warneke (student 203018141)*, hereby declare that the *treatise* for *Masters Degree in Music (Music Technology)* is my own work and that it has not previously been submitted for assessment or completion of any postgraduate qualification to another University or for another qualification.

Andrew Warneke

Official use:

In accordance with Rule G4.6.3,

4.6.3 A treatise/dissertation/thesis must be accompanied by a written declaration on the part of the candidate to the effect that it is his/her own work and that it has not previously been submitted for assessment to another University or for another qualification. However, material from publications by the candidate may be embodied in a treatise/dissertation/thesis.

ACKNOWLEDGEMENTS

- My supervisors, Dr. Rudi Bower and Mr. Mark Brand, for their support, encouragement and input through the course of this project

- My wife, Claire, for her patience with me and for not giving up hope that this treatise would get completed

- My family and friends for continued interest in my work

- The people who participated in the listening tests – without your involvement this project could not have happened

- The staff at Nelson Mandela Metropolitan University's School of Music for making facilities and equipment available to me when I needed them and for allowing me to complete this research

- Greg Stokell for allowing me leave from work to make headway at key times in the development of this study

- The Lord for giving me the will to persevere with this project and make it to the end

CONTENTS

TABLE OF FIGURES

TABLES

ABSTRACT

This study sought to determine whether or not it may be possible to evoke ‘the musical effect' – the emotional response perceived by music listeners – using white noise as a sound-source and real-time digital signal processing techniques. This information was considered to be valuable as in a world driven by technological progress the potential use of new or different technologies in creating music could lead to the development of new methods of – and tools for – composition and performance.

More specifically this research asked the question 'what is music?' and investigated how humans – both trained musicians and untrained people – perceive it. The elements of music were investigated for their affective strengths and new fields of research explored for insights into emotion identification in music. Thereafter the focus shifted into the realm of Digital Signal Processing. Common operations and techniques for signal manipulation were investigated and an understanding of the field as a whole was sought.

The culmination of these two separate, yet related, investigations was the design and implementation of a listening experiment conducted on adult subjects. They were asked to listen to various manipulated noise-signals and answer a questionnaire with regard to their perceptions of the audio material. The data from the listening experiment suggest that certain DSP techniques can evoke ‘the musical effect’. Various musical elements were represented via digital techniques and in many cases respondents reported perceptions which suggest that some effect was felt. The techniques implemented and musical elements represented were discussed, and possible applications for these techniques, both musical and non-musical, were explored. Areas for further research were discussed and include the implementation of even more DSP techniques, and also into garnering a more specific idea of the emotion perceived by respondents in response to the experiment material.

Key words: noise-control techniques; musical effect; elements of music

CHAPTER 1 - INTRODUCTION TO THE CURRENT RESEARCH

The purpose of this study is to determine whether or not there are digital signal processing techniques that show promise in their ability to evoke ‘the musical effect'[1] in humans, using only white noise as a signal source.

This chapter will deal with the background to the research topic selection; the rationale underpinning the study; the formulation of the problem statement; and the definitions of operational terms to be used in this treatise.

1.1 BACKGROUND TO THE RESEARCH TOPIC

The beginnings of the selection of this topic of study emerged from a discussion I had with a business executive who consults to various industrial plants. He approached me regarding the problem of excessive noise in the factories he consults to, as these companies are forced to spend a large sum of money annually to purchase hearing protection for employees who work in close proximity to machinery as the noise-levels are above the legal limit. He wondered whether there was not a way of using digital technology to reduce the noise-level, as acoustic control methods are rendered ineffectual due to the porous nature of the sound absorption materials (they become saturated with dust, and then only serve to reflect noise, instead of absorb it). Unfortunately, the technologies which could be used in this situation are still in their infancy, and research is still in progress as to how to expand Active Noise Control systems into a complex, visceral, sound-environment (Ruckman 2007). However, my discipline being music, this conversation did lead me on a path of creative thought regarding the use of noise in a creative manner. This study will purport to be the culmination of this creative thought process, together with a literature study, and an experimental study to test the effectiveness of the techniques uncovered.

[1] I ask that the reader grant me some latitude as regards the definition of the term "the musical effect", the precise definition of which will be a key focus of Chapter 3. In the meantime, I offer the following: "the musical effect" here refers to music's apparent ability to evoke an emotional response in a listener.

1.2 RATIONALE

The word "noise" is derived from the Latin word "nausea," meaning seasickness (Weil 2008). Thus, noise is cast in a negative light, as sound to eradicate from any musical situation, and is not usually given the opportunity to contribute to the sonic palette of music. However, some composers have tried to find a space for noise in the musical world. Noise is defined by Ambrose Bierce (1911:93) as: 'A stench in the ear. **Undomesticated music**. The chief product and authenticating sign of civilization' (emphasis mine). Bierce's thought, viewed one way, could lead to the conclusion that noise has music contained within it which could be brought-out.

So, if one were to think of noise from a creative viewpoint, as this treatise will attempt to do, noise could be viewed as a possible source of musical material, and the sources of noise as “instruments” to be manipulated, just as a pianist plays a piano. This gives rise to the desire to control noise to whatever ends the performer wishes. In order to “perform” on these new “instruments” the performer should have the ability to manipulate the noise in real-time.

The proposed treatise would set out to explore possible methods of real-time noise manipulation, with the goal of finding both practical and creative ways of manipulating noise to evoke the same effect in the listener (through noise) as common musical elements would. The precise nature of ‘the musical effect’ will be explored subsequent to this.

This research is significant as from my literature review it becomes evident that the manipulation of noise in real-time for creative purposes has received limited, and by no means exhaustive, attention previously. Thus the study would add to the body of musical knowledge. One possible contribution of this study could be to provide an inventory of practical techniques that could be used by composers to achieve new sonic colours in their works.

1.3 PROBLEM STATEMENT

1.3.1 Overview

This project aims to investigate, by means of a literature review and practical evaluation, whether or not digital sound processing techniques could show promise in their ability to evoke ‘the musical effect' in humans using only white noise as a signal source.

If there are some techniques which show good potential to achieve this end, a secondary objective could be to provide an inventory of digital techniques that have sufficient expressive capabilities to warrant their use for evoking ‘the musical effect’.

1.3.2 Research Question/Hypothesis

- Is it possible to evoke ‘the musical effect’ using real-time digital noise control techniques?
- What are the techniques that facilitate this end?

1.4 OBJECTIVES AND AIMS

1.4.1 Overall Objective

One of the objectives of an artist is to evoke an emotional response in his audience through his work. To this end, artists constantly search for new techniques of creativity. This research proposes to uncover more tools of musical expression, thereby opening the door to new creative possibilities for composers. The mere possibility that this study holds for the addition of new techniques of composition validates the necessity of the research. “What I’m most excited and passionate about is the idea of trying to encourage and to challenge composers and performers to try new things, new directions” - Peter Gelb (Freed 1998:3).

My contribution in conducting this research would be to provide an insight into the capability of noise to evoke ‘the musical effect’ in listeners when processed by digital means, and if the evidence is strong in this regard, to provide an inventory of implementable noise control techniques that could be used by composers to evoke ‘the musical effect’.

1.4.2 Specific Aims

The aim of this research is to ascertain whether ‘the musical effect’ can be achieved through the digital manipulation of a noise signal, and then to possibly identify a technique or various techniques that can be used to control noise in real-time to evoke ‘the musical effect’. These techniques should be practically reproducible with some training in their implementation.

The four main questions the study hopes to address are:

1. Is it possible to evoke 'the musical effect' using real-time noise control methods?

2. Which musical elements are the most evocative of 'the musical effect' when implemented via a noise-control technique?

3. Which properties of digital processing methods are best suited to the purpose of the research?

4. What are the possible applications of these techniques in both the music world, and other spheres?

1.5 OPERATIONAL DEFINITIONS

Music: Organised sound that evokes an emotional response in the listener (see Chapter 3).

Noise: Sound that is characterized by its non periodic (random), continuous structure (Chang 1999; Russ 1996; Roads et al. 1997).

Digital: "Representing data as numbers: processing, storing, transmitting, representing, or displaying data in the form of numerical digits, as in a digital computer." (Encarta: "digital")

Analogue: "Using physical representation: relating to a system or device that represents data variation by a measurable physical quality." (Encarta: "analogue")

Real-time: "As fast as required. A real-time system must respond to a signal, event, or request fast enough to satisfy some requirement." (PC Mag: "Real-time")

For the purpose of this study, we will accept "real-time" to be t + 25 milliseconds (where t is the time of operation).

Real-time may also refer to an operation which was recorded in real-time and played back to an audience, as the recording was made in real-time, and the audience experiences the same.

Offline Processing: Processing that occurs after a recording has been made i.e. not in real-time.

Musical Training: In this study, a person will be deemed to have musical training if he/she is enrolled in a tertiary music program.

Musical effect: An emotional response that is evoked by music in the listener (see Chapter 3).

1.6 CHAPTER OUTLINE

Chapter 1

This chapter functions as a general introduction to the project. As such, the background to the research topic selection is discussed, which leads to the development of a rationale. The problem statement is formed and research questions are then discussed. Operational terms to be used are given definition so as to clarify certain meanings which may be ambiguous or unknown to certain readers. The chapter concludes with a chapter outline.

Chapter 2

Chapter 2 deals with the design of the research to be undertaken, and an explanation of the methodological considerations is given. Included in this discussion would be sections on ethical issues involved in this project, the reliability and validity of the research, the techniques of data collection, and an explanation of the data analysis techniques to be utilized. Sampling methods and measuring instruments will also be included here. The chapter ends with a brief summary.

Chapter 3

This chapter makes up one half of the literature review that was undertaken in preparation for the experimental phase of the project. This portion forms an in-depth investigation into the meaning of music and why humans perceive it and are affected by it and a description of musical elements that could be used in evoking ‘the musical effect’.

Chapter 4

This chapter contains the second half of the literature review. A review of pertinent literature in the fields of noise, noise control, digital signal processing, and their real-time applications is given.

The literature review was broken into two separate sections as two fields are discussed. These could be seen as two axes on a graph. With the main aim of the research in mind, my hope was to find certain points of intersection of these two axes as these could point toward possibilities in terms of achieving the research goal.

Chapter 5

Chapter 5 describes the practical implementation of the noise control techniques unearthed in Chapter 4 in attempting to evoke the musical elements discussed in Chapter 3. The specific list of musical elements which will be reproduced by means of noise control techniques is discussed, and the selection of noise control techniques to be used is specified. Sample selection, questionnaire design, and design of the experimental environment are then dealt with.

Chapter 6

This chapter presents tabulations and analyses generated from the data collected from the experiments performed.

Chapter 7

Chapter 7 is a discussion of the conclusions drawn with regard to analysed data and a list of possible implications of the research project. Also, where possible, suggestions for possible further research will be made.

CHAPTER 2 - RESEARCH DESIGN

2.1 <u>INTRODUCTION</u>

The research design of this study consists firstly of a literature review separated into two distinct areas. The first is on the topics of music and the musical effect, and the second, Digital Signal Processing techniques and noise. The decision to split the literature review into two parts stems from the idea that this study must deal with two distinctly different areas of music technology – the human perception of music and digital technology – and then try to bring them together in the experiment in order to test the research question. Every effort has been made to separate these two domains. However, at some points there is unavoidable overlap due to the recent progress in the field of music technology with specific reference to music-signal analysis.

The goal of the first section of the review is to understand ‘the musical effect', and then to form a list of factors – the elements of music – that contribute to this phenomenon. The second section will seek to explore noise and digital signal processing techniques in order to attempt to find methods from the digital domain for evoking the musical effect in the experiments to be conducted later in the study.

Once these areas have been thoroughly investigated, I will then undertake a process of attempting to find points of intersection between the musical elements that were found and the digital signal processing techniques discussed, in order to create possible compositional tools for evoking ‘the musical effect' using digital techniques of noise-manipulation in real-time.

Once these proposed tests have been created, a questionnaire and experiment will be formulated and performed on members of the public (specifically adults) of undefined levels of musical awareness. The experiment will take place in an environmentally controlled venue containing a suitable music reproduction system and comfortable seating for the respondent. This research will compare noise control effects on people of varying (unspecified to the researcher) levels of musical training. This is appropriate as music is very seldom performed to an audience who have all achieved the same degree of musical expertise. Thus, in trying to garner an impression of the effectiveness of the techniques uncovered I feel that a 'mixed' audience is beneficial. Respondents will complete a questionnaire as the tests are administered. The questionnaire will deal with the effectiveness of each digital signal processing technique in evoking ‘the musical effect'.

Once the experiment is completed the data from the questionnaires will be collated and analysed in order to ascertain the effectiveness of the various techniques utilized in the experiment in evoking ‘the musical effect'. A conclusion will then be drawn as to whether there are any techniques uncovered in this study that may have merit as tools for music composition.

From the above description of the study one can see that the research methodologies to be used in the course of this process will come mostly from the positivistic paradigm which combines literature review and quantitative methods of data collection (Babbie and Mouton 2005). Common pitfalls of positivist research such as unfounded generalisation and the unprecedented ascription of causal relationships will be avoided. The study has an exploratory element in that the study seeks to uncover new potential compositional tools, and because the study will create new data, a 'primary data design' will be presented through the data collected from completed questionnaires.

Due to the necessity for this study to produce statistical evidence of the effectiveness of the noise control techniques in evoking ‘the musical effect' at its conclusion, a quantitative approach will be adopted (Adler 2008:432; Babbie and Mouton 1998:49). This, together with the subject matter at hand, has led to the decision to use closed-ended questions in the questionnaire. This will allow me to present language that is understandable to the respondents, leading to more clearly defined categories of answers, rather than asking respondents to explain in their own words phenomena which may not be easy to put into words.

2.2 ETHICS AND HUMAN SUBJECTS ISSUES

Research carried out on human subjects must be approved prior to data collection by a committee registered with the National Health Research Ethics Council of South Africa, which has been established in terms of the National Health Act 2003 (Act No. 61 of 2003). However, because I have chosen to use only adult respondents – who are not considered a vulnerable group – in the experiments, and the nature of the experiments did not put respondents or myself at risk, ethics clearance did not need to be obtained from any of the NMMU ethics committees.

According to Creswell (2009), researchers must anticipate and try to pre-empt ethical issues that may arise throughout the research process. Researchers are responsible for protecting research participants and developing trust with them, and conducting research with integrity and avoiding misconduct and impropriety. To ensure that the research was conducted with understanding of and due regard to acceptable ethical standards, the ethical considerations outlined in the following sections were given credence.

2.2.1 Non maleficence

Israel and Hay (2006) define the principle of non maleficence in research as the minimisation of any risk or harm which may accrue to participants as a result of the research process being carried out. The present research poses little risk for participants, since the questionnaire is to be administered without the respondent writing his/her name on the questionnaire. This excluded the possibility of identification of respondents and ensured that confidentiality would be maintained. According to Mouton (2001), a further requirement in terms of non maleficence is that the research results are freely and openly disseminated. The final research reports of postgraduate research projects are made available on the NMMU intranet. In addition to this, I will make my contact details known to the respondents at the experiment so as to allow them to obtain the final product from me after it is completed. This means that results will be available to participants once the research has been carried to completion.

2.2.2 Beneficence

Beneficence refers to an expectation that the researcher enhances the well-being of the participants or the society as a whole, through the conduct of the research process (Israel and Hay 2006). Walsh (2001) emphasises that research should not be carried out purely for the purposes of enhancing the career or reputation of the researcher, but should provide, on balance of cost and benefit, some value to the participants or the broader society in which they are found.

This research will not necessarily add direct value to the participants who choose to respond to the questionnaire. However, if a new musical technique is uncovered through the findings of the tests, society would benefit in that composers may be enabled to create new music to be enjoyed by the public.

2.2.3 Social justice

The principle of autonomy was adhered to since it was assumed that individuals who chose to participate did so of their own volition. The questionnaire was preceded by an introduction explaining the details of the study to ensure that consent was given by individuals who were duly informed of the relevant facts. Individuals were informed of their right to withdraw from the study at any stage.

2.2.4 Fabrication, falsification and plagiarism

The concept of research misconduct is made up of three components, namely, fabrication, falsification and plagiarism. Fabrication is defined as "making up data or results and recording or reporting them" (Israel and Hay 2006:113). Falsification is changing or manipulating research materials, equipment or processes, or changing or leaving out data or results in a way that causes the research not to be accurately reported. Plagiarism is the act of using another person's ideas, processes, results, or words without giving credit to that person in the appropriate manner. In order to avoid plagiarism the researcher must ascribe appropriate authorship to the writer of the original idea.

I understand that plagiarism, falsification and fabrication constitute research misconduct and have adhered to ethical research conduct in carrying out the present research. The literature study was conducted with full understanding of the consequences of plagiarism and sources are fully acknowledged and referenced.

2.3 RELIABILITY AND VALIDITY OF THE RESEARCH

2.3.1 Reliability of research

According to Mitchell (2010:143) the reliability of a study is ensured through the development of a system of testing and analysis that gets the same results – or as close as possible to – every time. A study cannot be valid unless it is reliable. The reliability of this research will be established through the use of a sufficiently large sample of mixed gender and cultural background; and a guided statistical analysis of the results of the questionnaire. The experiments will also be conducted using an equipment and venue set-up which is readily available and easily replicable. Factors such as volume-level will be monitored so as to ensure that the experiments can be performed the same way every time, and thus, the respondents' experience should be as close to identical as is possible.

2.3.2 Validity of research

Hammersly (in Alasuutari et al. 2008:43,44) points out that the 3 criteria in quantitative research that should be used for assessing validity in a study are: measurement (measures must be reliable and valid), generalization (is the test likely to yield results true for the larger population?) and control of variables to exclude random error. He says that causal validity is a measure of the validity of predictive hypotheses, or answers the question: "Is there a good reason to conduct this study?"

He also notes the differences between internal and external validity as being that internal validity is a measure of the study’s systems of measurement and causal validity, whilst external validity speaks of generalizability.

There are three sorts of knowledge claim that can arise from a valid study: descriptive, explanatory, and theoretical. The conclusion of any study should fall into one of these three areas.

This particular study hopes to make, at its conclusion, a primarily descriptive knowledge claim in that its goal is to show that certain noise control techniques could be used to evoke ‘the musical effect'. The test results will describe how subjects responded to the test stimuli and then theorize as to how these methods could be used as compositional tools. In assessing the validity of descriptive claims we must ensure that features attributed to the phenomena to be described are actually part of the phenomena, and whether they are possessed to the degrees indicated (Alasuutari et al. 2008:44). Through a review of available literature and careful experiment design, I feel confident in achieving an accurate and valid description of all phenomena included in this study.

The measurement of the effectiveness of the techniques is reliant upon the answers of the sample group. The variable which must be considered here is sample size – a sufficiently large sample group should be drawn in order to generate an appropriate amount of data from which to draw conclusions. The measurement of effectiveness should prove internally valid, as long as a sufficiently large sample is drawn. Validity will be further aided by alerting the subjects to the fact that some parts of the experiment have been purposefully unaffected by the control techniques, in an attempt to ensure that the subjects answer the questionnaire with the proper integrity.

Generalizability is limited to (at best) the adult population of Port Elizabeth, South Africa. Further studies could be undertaken in the future to determine the true generalizability of the results obtained here.

This research deals with the specific use of noise control techniques to evoke the musical effect. The effectiveness of the techniques will be limited to their effectiveness in evoking only the musical effect as discussed later in this work. This research question does not include the evaluation of any noise control technique that does not work in real-time, and the study will test only the effectiveness of the techniques in evoking the musical effect. The playability of the technique is not under scrutiny.

The validity of a study is not a concept that can be quantified or measured. However, based on a solid approach to experiment design I feel confident that this study is reliable, and internally and externally valid.

2.4 VARIABLES

Nardi (2006:44) defines a variable as being a measurable concept. The 'dependent variables' of a study answer for why the variable varies – the variability being dependent on various causes. These causes or explanations are the 'independent variables' in the study. Whether a variable is dependent or independent is a function of the particular study. Adler (2008:187) concurs with this view when he states that in experimental research the independent variable is not to be measured, but is rather 'introduced, manipulated or controlled' by the researcher.

In this study the variables will be:

Independent: musical elements presented via noise control techniques in listening experiment; age of respondent; gender of respondent; time of day of testing

Dependent: response (measured on questionnaire) to listening experiment

2.5 DATA COLLECTION

2.5.1 Sources of Data

- A literature review of sources in fields relevant to the research at hand

- Questionnaires completed by respondents in response to the experiments performed

2.5.2 Population and Study Sample

Unit of analysis: The unit of analysis is the effectiveness of each technique, on each subject, in evoking the musical effect.

Target Population/Sample frame: The target population for the study will be male and female adults of no particular affiliation to me or to one another. All will, however, be residents of Port Elizabeth, South Africa.

2.5.3 Sample Size and Selection of Sample

The sampling method to be employed in this study is non-probability convenience sampling, which is used in much exploratory research and involves the researcher selecting a sample group based on his judgement of what units will facilitate a fruitful investigation of the subject matter and then utilizing available subjects from this group (often used in university studies). This method applies because the experiment will be conducted in a time where not everybody who is asked to attend may be able to be present, and thus allowance must be made for this factor on the days of the experiment. Convenience sampling has been shown to yield results that are 'wonderfully provocative and plausible' (Adler 2008:121). In this case I have selected a sample of adults from Port Elizabeth as I believe this to be an effective group on which to test the effectiveness of the tools uncovered in the literature review. I will endeavour to test in excess of 30 respondents, as according to Hogg and Tanis (2005) and Kish (1965), the size of a sample needed to provide meaningful results should be equivalent to or greater than 30 units. Although the 30 unit 'Rule of Thumb' has been argued by other writers such as Cohen (1990), the limited scope of this project, as well as the descriptive analysis to be applied to the data generated by the experiments, allows for a simple approach to sample size. Descriptive statistics are used when the goal of research is to summarize a sample and describe the main features of the data collected rather than to learn about the population the data represents (Mann 1995; Dodge 2003). This study sets out to explore the potential of DSP techniques used to manipulate a noise-signal to evoke the musical effect, and does not purport to claim any universal truth on the subject matter at hand. Thus further research could certainly be carried out using a much larger sample in order to make claims which are representative of a larger portion of the population, but with this goal in mind a sample size of 30 will suffice. Further to this, there is evidence of other studies, including ones by international organisations like OECD (Business Tendency Survey Handbook 2003), which employ the 30 unit sample size idea in their work.

There are three main sampling-related errors to be avoided in any study: coverage error (a difference between the sampling frame and target population); non-response error (respondents refuse/fail to answer questions) and sampling error (the selected sample does not represent the wider population) (Adler 2008:106).

Coverage error can hardly be avoided when dealing with a study that relates to anyone who is able to perceive ‘the musical effect', as this is essentially true of all of humankind – as shown in the literature study. As such, one can only hope that the respondents are diverse enough in terms of culture and background that they represent a wide portion of music listeners.

Non-response error should not be a factor in this study as the questions on the questionnaires will not seek to elicit private or sensitive information from respondents and as such respondents should have no problem answering the questions.

Sampling error will be avoided in the use of a gender-mixed and culturally diverse sample group – the general adult population of Port Elizabeth.

2.5.4 Collection of Data

Data collection will take place through the implementation of a questionnaire to be answered by the sample group, in response to their individual perceptions regarding the material presented to them in a controlled environment.

2.6 DATA ANALYSIS STRATEGIES

Data from the questionnaires will be taken in directly after the experiment, and captured, tabulated and analysed. Responses from the questionnaires will be kept separated according to quota groups (if these are included) in order to allow for greater variety of analysis and conclusion.

The data to be analysed will be the responses of respondents on the questionnaires detailing their perceptions of the emotional content of the various manipulated noise signals played to them in the experiments. The aim of this analysis is to ascertain whether or not the manipulated noise signals have a measure of evocative power. In other words, could any of the manipulated noise signals be used as a musical device to evoke emotion? The analysis will also seek out any pattern in responses which could be traced back to the influence of an independent variable, and if these patterns are found, the effects of the independent variable will be dealt with.

2.6.1 Measuring Instruments

The measuring instrument used in the listening experiment portion of this study will be a self-administered questionnaire, however I will be present at the time of the experiment as the tests will be conducted by myself.

Benefits of this style of research include factors such as the researcher being available to encourage reluctant respondents to participate and provide additional instruction/explanation of the questionnaire. However, at the same time, these benefits may also cause respondents to alter their responses to questions, especially when discussing more sensitive topics – not really a factor in this study (Alasuutari et al. 2008:314).

In order to ensure the greatest possible degree of transfer of intent, the questions asked on the questionnaire will be worded in language that is understandable and clear to the respondents. This will be achieved through the use of simple words, the avoidance of ambiguity, and the asking of only one question at-a-time. Specification error will be avoided through the use of questions targeted directly at the objectives of the research question (discussed in Chapter 1) (Alasuutari et al. 2008:316). Some other items that will be avoided in order to ensure clarity and reliability in response to questions are that questions asking for opinion or agree/disagree will be made into personal statements of the respondents; negative statements will be excluded from questions and statement direction will be mixed so that questions do not all lead to positive responses (Nardi 2006:76-78).

2.7 CHAPTER SUMMARY

This chapter includes discussions of data sources to be used in the study, population and study sample to be used, the sample size and method of selection, methods for data collection and analysis, and the measurement instruments to be used in order to gauge listener response in the experiment.

CHAPTER 3 - LITERATURE REVIEW PART 1

3.1 MUSIC AND 'THE MUSICAL EFFECT'

The primary goal of this project is to answer the question: "Can one evoke 'the musical effect' from 'noise' in real-time?" In order to do so, an understanding of music, 'the musical effect' (the emotional response that music evokes), and common musical devices or musical elements must first be established. In this chapter an exploration into these areas will be undertaken in order to identify a list of elements of music to try to emulate using noise-manipulation techniques.

The first question which presents itself then, is 'What is Music?' No culture in recorded history has ever lacked music (Levitin 2006:6), and much research has been undertaken over many years in the effort to produce a definitive answer to this question. There are many points of agreement and disagreement between the various texts available for reference, and thus I have chosen to present some of the ideas that seem most pertinent to the study at hand.

One prevalent viewpoint sees music as sounds, systems, and rules. Varese says that 'music is organised sound' (Goldman 1961:133). This sounds like the simple definition that captures 'music' conclusively, yet Deutsch (1982) and Chase (2006) would argue that patients with 'amusia' – the psychological condition of lack of musical ability and understanding – can sometimes sing the tune of a song, but in their research the performance of the tune was not perceived as 'musical'. They found that this was because there was no spontaneity involved in the process of the creation of the sound. So 'music' could be more than just 'organised sound'. Could it be sound organised in such a way that it evokes an emotional response in the listener? One possible reason that this could be true is that music is thought to have evolved as a means of emotional communication (Chase 2006:5). It follows then that a musician should practice basic motor skills to the extent that they become automatic, in order to concentrate on the stimulation of affect whilst performing. So 'music' doesn't come from the motor skills or the physical facility of the performer, and therefore, the creator of music – the musician – must not only create the sound through his well-practiced motor skills, but also stimulate an emotional response in the listener. Something about his thinking or feeling must create this response. Xenakis (1971:181) agrees in saying that 'musical action demands reflection – the mind must exert some control over what is created'.

One of the traits that organised sound should possess to be called music then is that it should be able to be perceived emotionally, mentally and spiritually, that is, it should affect the human beings listening to it. Juslin and Sloboda (2001) suggests that an emotional experience of some sort is the main reason behind most people's enjoyment of music. Macone (1990:11) writes that 'to experience sounds as music is to experience something capable of being shared'. It 'represents pleasure, distraction, avoidance of thought, inability to confront serious intellectual issues, brute instinct, unreason and ultimate despair, a value system impervious to argument' (Macone 1990:173). These writers agree in their belief that music is a social experience that is enjoyed by listeners due to its ability to affect the way we feel. Furthermore, they agree that 'the sound element in music is a powerful and mysterious agent' that 'has an expressive power' and is a 'potent and primitive force' (Copland 1963:2). So we understand that, as humans, we are instinctively drawn to music for its effect on us, and that we also do not fully understand it. It is an experience enjoyed by most and seriously understood by few (Macone 1990). Copland proposed that 'our musical instinct is not all that elementary; it is, in fact, one of the prime puzzles of consciousness' (1963:18).

Macone (1990:10) defines music as 'the idea behind or beyond the sound' and adds that 'music sums up the art and experience of listening and communicating in sound' (preface). If 'music' is the idea and it is experienced as a form of communication, then successful music must convey an idea to or evoke an effect in the listener. The meaning of the idea may be communicated clearly or may be perceived as an emotion. The feeling/idea does not need to be the same for every audience member. Copland wrote that music expresses moods, and that the meaning conveyed by a piece of music may be visceral (1963:4), and Meyer (1956:18) says that music needs only to evoke a feeling in the listener to have accomplished its task, not necessarily a specific image. In fact, strictly speaking, the widely held notion that music is a language comes under serious fire with the idea of Sheperd and Wicke (1997) that music may be a system of signs, but it does not necessarily signify definite objects, and thus cannot be considered a language.

Mazzola (2002:4) defines music as 'a system of signs composed of complex forms which may be represented as physical sounds, and which may in this way, mediate between mental and psychic contents', and takes the view that music expresses the musical state of its creator, and emotionally affects the listener because it is involved in 4 artistic and scientific actions: production; perception; communication and documentation. He sees these 4 actions as intertwined, and as congruous with the major philosophical concerns of life. These authors seem to be saying that people are drawn to music because they can relate to it, even if they are unaware of its technical value (Mazzola 2002:284), because the musical experience simulates (through sound) the daily human experience.

So, from all of the above, what can we say is 'music'? Could we say that 'music' is 'organised sound that evokes an emotional response in its listeners'? I will accept this definition as it is congruent with the views of other authors and serves the purposes of the current research well. It follows from this definition that ‘the musical effect’ could be the emotional response which is evoked by music. The above definitions are by no means absolute or universal. Music can be defined in numerous ways and there may be other musical ‘effects’ that could be described. However, with the goals of the current research in mind these definitions are appropriate and more than adequate as they speak directly to the research question.

With this definition of music in mind, one should now be led to ask the question: 'How does one come to experience this affective organised sound?'

Deutsch's studies cited above seem to place the responsibility for creating affective music on the performer of the music, and Levitin (2006:204) says that the essence of good musical performance is the ability to convey emotion.

Whilst these statements may be true, one must say that they can only be partly true, as there are many factors other than a good performer which are necessary for an affective musical experience. Nattiez (1990:ix) says that the musical work is not merely the score; and it is not just its structures. Rather, for him 'the work' also constitutes the procedures that have produced it (act of composition), and the acts of interpretation (from the performer and listener) and perception (by the listener). So music is the composition (on paper or spontaneous), a great artist interpreting this composition, and the audience's response to it.

Nattiez is not alone in attributing some responsibility in the musical affect to the listener. Many authors argue that one may only truly appreciate music through the serious study of its component parts. The question that must be answered here though is “What are these 'parts' to the average listener?”

Copland tries to make a case for all listeners of music having the responsibility to study music more rigorously in order to enjoy it more. He says:

> *It is very important for all of us to become more alive to music on its sheerly musical plane. After all, an actual musical material is being used. The intelligent listener must be prepared to increase his awareness of the musical material and what happens to it. He must hear the melodies, the rhythms, the harmonies, the tone colours in a more conscious fashion (Copland 1963:5).*

This view is one that is not only held by Copland. However, as discussed later in this chapter, it is not a view that I feel represents the truth of the issue at hand.

Macone (1990:12, 15) says that music is pleasing to listen to when the elements of music are clearly defined, and that the training of the listener dictates his perceptual ability. Meyer (1956) agrees that the listener's response to music is dependent on his musical training, and says that, in general, the further an element is taken from normative levels, the more likely it is to evoke an emotional response. (Could this be why 'pop' music is so extreme in its use of musical elements?) He believes that the content of a listener's mind will affect the meaning he draws from the music (Meyer 1956:16).

Mazzola (2002) has said that the most prevalent belief amongst the public is that music is emotions, showing that listeners expect to be affected by music in an emotional way, and that without this perception of emotions by the listener (though the level of perception may vary from listener to listener), one cannot truly say that a musical work has been created.

This leads to the idea that the schools of thought can be combined as follows: the creator/producer of a piece of music must understand how to use his music to affect his audience in the desired way. This understanding can be achieved through a thorough study of music. It is my belief that all listeners possess an innate ability to be deeply affected by music, as human-beings all seem to be born with an appreciation for music (Chase 2006:21). The study of music may allow the listener to understand the intent of the composer/performer in terms of the affect they hoped to achieve, but that should not dictate one's honest response to the musical material being presented, and does not mean that one who has no technical musical understanding cannot appreciate music and be affected by it. Recent research agrees with this sentiment and Levitin (2006) shows that the listener's first impression is of overall sound of the music, and that the average listener never gets past that stage. Thus the majority of people are affected by the sonic whole, rather than separate musical elements. So for most listeners, the effect of music is felt through all elements together, and although they probably cannot specify what each one did, this does not stop them from experiencing the musical effect. Bigand and Veillard (2005:1113) shows that emotional responses to music are very stable within and between participants, and are weakly influenced by musical expertise and excerpt duration.

So it is clear that musical elements, working in conjunction with one another, evoke what we call the musical effect and move the listener emotionally. In trying to understand how it is that music affects the emotions, I feel one should first seek to understand – or address – the process of how humans most naturally experience music. It has become the norm to talk about the experience of music-listening from the viewpoint of a trained musician, as we have developed a complex system of labels – such as 'melody', 'harmony' and 'rhythm' (and many others) – by which to communicate.

This level of discourse becomes problematic, however, when one begins to think about how the untrained and uninitiated person hears a piece of music. I refer here to a listener who may never have heard of 'melody' or 'rhythm', or even a child who has never heard music. McLean (1981) asserts that Western music beyond the common practice period has set up an elaborate pitch-rhythmic-notational system. He argues that this system, in its extreme development, has alienated us from our own basic human body gestures, rhythms, and melodies and that this has effectively amputated humankind from its own sound art. He proceeds to call the systematisation of music a censoring device, an artificial creation which is robbing people of the natural experience of music. When one considers the sense of wonder with which most children respond to music one could be led to believe that there is truth to this – especially with regard to the idea that with age and musical understanding there seems to be a decrease in the 'naturalness' with which these same people are able to respond – an idea that warrants further investigation, but will not be tackled in depth in this work. They have, since childhood, been taught musical 'right from wrong', and as such are unable to lose themselves and hear the way they did when there were no limitations and systems imposed on them. So what started out as a necessary set of mnemonic devices for musical practice gradually started determining what that practise should consist of, and as such, what 'real music' should sound like. Surely this also affects how we hear the music, and as such, how we are affected by it.

So what would be left of music if one were to discard the systematised, textbook-orientated musical jargon? How would one hear without any of these academic crutches to lean on? How would one talk about the music experience without making use of the standard 'buzzwords'?

It is a common practice in linguistics to divide the sign into the signifier and signified. The first to do so was Saussure, whose definitions for these components or 'faces' planted him firmly within the territory of psychology. The basis for his concept – called 'structuralism' - is that all phenomena of human life are not intelligible except through their network of relationships, an idea which gives rise to the sign and the system (or 'structure') in which the sign is embedded – the primary concepts. Hence, a sign - for example a word - obtains meaning only in relation to or in contrast with other signs in a system of signs. His basic idea led to a signifier/signified distinction related to a split between the "sound-image" and the "concept" (Saussure 1983). He points out that the linguistic concept is neither the sound nor the thought, but rather, the whole link that unites sound and idea, signifier and signified. Saussure: "A sign is not a link between a thing and a name, but between a concept and a sound pattern" (1983:66). In the English language the signifier "dog" signifies an animal of the canine species. It is quite specific as it is clear to speakers of this language – or structure – that the sound-pattern denotes a canine (each individual listener's response on an emotional level to the signifier may vary based on various things like past experience and culture but all would agree that “dog” refers to a dog). But it is also completely arbitrary in that the only reason we understand the meaning of this sound-pattern is that this is the meaning that we are taught from birth. Thus, an essential feature of Saussure's linguistic sign is that, being intrinsically arbitrary, it can be identified only by contrast with coexisting signs of the same nature which combine to form a structured system.

Saussure went on to provide a basis for the expansion of his science of signs to all disciplines: He called this new science 'semiology'. It was to investigate the nature of signs and the laws governing them. Roland Barthes helped greatly in founding the science of semiology, seeking to apply structuralism to the "myths" he saw in the everyday world: media, fashion, art, photography, architecture, and especially literature. For Barthes, "myth is a system of communication." It is a "message," a "mode of signification," a "form" (Barthes 1973:109). Barthes extends Saussure's structuralism and applies it to myth as follows:

Myth is a peculiar system, in that it is constructed from a semiological chain which existed before it: it is a second-order semiological system. That which is a sign (namely the associative total of a concept and an image) in the first system becomes a mere signifier in the second. We must here recall that the materials of mythical speech (the language itself, photography, painting, posters, rituals, objects, etc.), however different at the start, are reduced to a pure signifying function as soon as they are caught by myth (1973:114).

Because semiology relates to all phenomena of life, it makes sense that it must be applicable to music too. As discussed earlier in this section we have, in music, created a system of what Saussure would call signs which have meaning by virtue of what we are told they signify. In music the signifier/signified relationship may be somewhat more complex than just sign related to meaning. This is because the system we have developed has been conceived in a multi-layered arrangement where a particular sign often just leads to another sign, or could lead to multiple interpretations. For example, if one were to use 'sounds' as a sign, then there exist myriad possibilities for the meaning of that sign. One could interpret 'sounds' to be an absolute in which case the sign references the complete sound, but this is seldom the case in music. We often use musical terms to refer to other terms which have meaning as individual entities within the system and, as such, musical semiotics includes much vagueness.

In addition to this problem, we are confronted with the added difficulty of cultural bias both in terms of the signifier and the signified. Phillip Tagg (1999:17) makes reference to some points at which music seems to exhibit universal features. These are features of music with parallels to the workings of the human body and human gestural traits. He calls these features 'bioacoustic'. The basic ‘bioacoustic universals’ of musical code can be summarised as the following relationships:

1. Between musical tempo (pulse) and heartbeat (pulse) or the speed of breathing, walking, running and other bodily movement. This means that no-one can musically sleep in a hurry, or stand still while running.

2. Between musical loudness and timbre (attack, envelope, decay, transients) and certain types of physical activity. It is difficult to make gentle or ‘caressing’ kinds of musical statement by striking hard objects sharply, and it is counterproductive to yell jerky lullabies at breakneck speed and that no-one uses legato phrasing or soft, rounded timbres for hunting or war situations.

3. Between speed and loudness of tone beats and the acoustic setting. Quick, quiet tone beats are indiscernible if there is a lot of reverberation and slow, long, loud ones are difficult to produce and sustain acoustically if there is little or no reverberation.

4. Between musical phrase lengths and the capacity of the human lung. Few people can sing or blow and breathe in at the same time. It also implies that musical phrases should last between two and ten seconds.

He points out that although these relationships seem universal, the emotion with which the phenomena are perceived may still vary widely across cultural groups. Apart from cultural difference he also notes that the way in which these relationships are combined within a piece of music also determines the perception of emotional meaning. (Although an exploratory study of these combinations seems interesting and could prove fruitful, they fall beyond the scope of the current research, but could be an opportunity for further investigation.) Thus the structures by which we process the significance of various signifiers are developed on the basis of one's cultural background, and as such, we cannot say (though it is frequently heard), that music is the universal language (Tagg 1999:18). As Peter Kivy points out in his example of the Bach-bird’s song which brought to his mind the Bach Organ Fugue in C:

While my musical consciousness is busy fitting the Bach-bird’s song into a possible Western notation, based on major and minor seconds, the Indian’s musical psyche is just as busy *accommodating it to the world of microtones. Does he hear the cheerfulness of the Bach-bird's song? Why should he? He doesn't hear the same 'music' in it that I do. I should no more expect him to hear the cheerfulness in the Bach-bird's song than the cheerfulness of Bach's fugue. For he would need my musical culture to hear both of them; and that, by hypothesis, he does not have* (Kivy 1980:92).

Kivy's point is that he reconciles the meaning of what he hears in the light of the fact that he hears everything based on his Western musical background, whilst someone who learned music in a different cultural setting may perceive something entirely different due to his prior experience in his own culture.

Kivy's observations seem to confirm Tagg's assertion regarding the culturally-based perception of music, and also Saussure's idea regarding language. This is, according to Saussure, that every meaningful sound is immediately connected with a certain concept. Moreover, Saussure highlights the non-necessary connection between the mental image of an object and a particular word in a spoken language. In other words, he posits that the link between sound and meaning is purely a social convention, one based on the particular society in which one is speaking. This is not to say that one individual within a particular language community can call an object whatever he or she decides to call it. Saussure's point is that once you are part of a particular language community, the arbitrary origin of the connection becomes stable and even something to which I, as an individual member of that community, must submit.

The above observations point to the idea that within our Western-musical culture there must be a set of norms or conventional meanings for musical signs. This becomes a complex problem though, if one considers the usual means we have adopted – possibly by default – for assigning significance to musical sound. A particular group of notes may be treated in an almost infinite number of ways to produce different 'meanings' to the listener. Even then, listeners may vary in terms of their perception of musical meaning and thus the affective characteristics of this group of notes. Trained musicians may hear this group of notes and be able to make judgements based on our system of labels – categorising the notes into scale, modality, rhythm, and dynamic-level to come-up with an interpretation of the intended meaning. But in doing so we return to the initial problem. People do not hear a spoken word and then process it by category and then arrive at its meaning, rather, we hear and perceive. We do not perceive only the image of the signified, but also an emotional tone. It follows then that the average human-being should not hear music and then categorise it by feature before an idea of its significance begins to form in his/her mind. Children certainly do not seem to do so, and yet are affected by the music they hear. This all goes to say that I feel that in trained musicians there may be – through the creation of this system of musical labels - a third level to the signifier/signified concept in music. We have the music (signifier); the labelling process (intermediate level); and then a perceived meaning (signified). We know, however, that this cannot be how we are meant to experience music naturally. We should have some sort of signifier/signified relationship with no systematised processing in between. Phillip Tagg (1999) certainly calls for a renewal in terms of how musicians describe musical sounds in order to make them understandable and relevant to the experience of the common man. He cites the work of sociologists Frith and Goodwin (1990) who call for help from musicians in their attempts to understand the impact of pop and rock music on modern society. What they found was missing from both musical and sociological sides of the issue was the ability to connect music as sounds with the society in which it exists, which influences it, and which it influences. This means discovering which sounds mean what to whom, and in which context. The problem they had with musicians is that they found that the terminologies employed by musicians had very little relevance to the people and societies they were attempting to study. Tagg (1999:23) goes on to show that musical jargon frequently used by musicians are not necessarily signifiers as, to qualify as a signifier, the sign must have meaning from both the transmitting and receiving end of the communication stream. He calls for a system of musical signs which can be understood by listeners and be utilized by musicians.

All of the above leads me to believe that there must be another plane on which we could hear music, and on which musicologists and sociologists should seek to understand music and its effects on individuals and society-at-large. This 'other' plane is in some ways simpler, yet by virtue of its naturalness, in my view, higher. This is, I must concede, a point of conflict between different musicological viewpoints. Where some regard natural music perception to be the highest level of musical awareness and most relevant to the world today (see above), others would regard it as the most basic and thus lowest plane. In his seminal article entitled 'How We Listen' (from 1939) Copland divides music listening into three planes. He named these: (1) the sensuous plane, (2) the expressive plane, (3) the sheerly musical plane. He explains that anyone who is untrained in music is limited to listening to music on the 'sensuous' plane. That is, listening to music for the sheer pleasure of the musical sound and the way in which it affects the emotions. He says on page 1 of the essay:

It is the plane on which we hear music without thinking, without considering it in any way. One turns on the radio while doing something else and absentmindedly bathes in the sound. A kind of brainless but attractive state of mind is engendered by the mere sound appeal of the music.

Copland goes on to talk about how he feels that listening to music on only this 'simple' plane is an abuse – that although the sound-appeal of music is potent and attractive, we should not allow ourselves to listen only on this level as the mere sound is not the whole story. He encourages people to learn to hear the musical and compositional devices employed in the music and engage on the 'higher' planes of listening (1939:2). As much as this view has merit from an academic standpoint, I cannot agree with it as the ordinary music listener in today's consumerist culture seems to desire to be affected by music, but has little drive to delve into an academic study of its integral parts. I am also not sure that the parts that Copland and his contemporaries' school of thought would break music into would even be appropriate in view of what is to follow.

If one were to discard all of the labels and simply 'experience' music I believe what remains is the sheer sound – an acoustical event comprised of many smaller parts which, when added together, produce an 'affective force' unlike any other in nature. What is important here is that this sound-event is not meant to be broken down into its individual components but rather understood as one idea as far as possible. I say 'as far as possible' in that the goal of the treatise is to uncover which elements of a musical sound affect the emotions of listeners. Because of this intent the various elements in the music must, at some point, be inspected and interrogated for their individual affective strengths which contribute to the affective strength of the whole musical sound. However, in order to preserve the naturalness of the process, whilst at the same time facilitating a means of analysis (and later, manipulation), we return to the need for a set of labels that define the individual elements of a musical sound. It may thus appear that this has become a circular argument but this is not the case. As argued above, I do not believe that the untrained music listener gives attention to – or even understands – music from the viewpoint of traditional labels such as melody, harmony and rhythm. I do, however, posit that the same listener could and does knowingly perceive a set of more 'global' features of a musical sound when he/she listens, and that these features are the ones which drive the listener's emotional response to the sound. These are features which are not to be found in music theory textbooks, but have been uncovered in the world of music technology, and specifically a field called 'Music Emotion Recognition' (to be investigated later in the chapter). It is inside this field that I feel that there lies a fresh insight into how we perceive music, and how we could endeavour to analyse musical sound from a more natural perspective in order to gain an understanding of how music affects our emotions, and whether or not the same effect could be achieved through the digital manipulation of a noise-signal – which is the goal of this work: first to ascertain as to whether or not the potential exists for noise to be digitally manipulated to evoke an emotional response, and if it does, to provide composers with the means to manipulate a non-musical medium such as noise to evoke the same effect that music does. Bearing in mind the idea that affective music is created through the meaningful and balanced use of the many musical elements in relationship to one another (Levitin 2006:18,75; Meyer 1956:126; Chase 2006), a study of these elements and the way they affect listeners must be undertaken.

3.1.1 Common Musical Elements (or) Components of Musical that Evoke the Musical Effect in Listeners

Although this section deals with the necessary understanding of the elements that form 'musical sound', I feel that a brief exploration into the cognitive processing that takes place when hearing musical sound would be beneficial. This is necessary because although the brain may not be an element of music, it is certainly the filter through which all the elements must pass, or, as Chase (2006:16) asserts, the brain is where music is really made. As Smith (1979:169) says:

Raw sound becomes musical sound only after it has been processed, as it were, through the grid of definite categories.

An understanding of how the brain works to process sound through this 'grid' would aid us in understanding the transition from 'sound' to 'music'. It should also provide some insight into the origin of our emotional response to music.

The first point to note is that the human brain processes speech, noise and music as different things (Deutsch 1982:462; Levitin 2006:174). In fact, it has been shown that the cerebellum reacts strongly to music, but not to noise (Levitin 2006:174) – an interesting observation in the light of the goal of this research. What if the noise is manipulated to mimic musical elements?

The basic elements of any sound are loudness, pitch, contour, duration, tempo, timbre, spatial location and reverberation. Our brains first decide if the sound is 'musical' in nature, and then organize these elements into meter, harmony and melody (Levitin 2006:14). The brain processes each element of the sound separately (using frameworks or ‘modules’ it has built), but in parallel (simultaneously), then reassembles what we perceive as the 'music', and what psychologists call a 'gestalt' (or unified whole) (Levitin 2006:76,89,116; Meyer 1956:87; Chase 2006:17). Meyer (1956:6) wrote:

Understanding is not a matter of perceiving single stimuli, or simple sound combinations in isolation, but is rather a matter of grouping stimuli into patterns and relating these patterns to one another.

In a similar vein, Levitin (2006:109) calls music a 'perceptual illusion' in that our brains impose order on a sequence of sounds that may have had no meaning prior to this processing. Another noteworthy point is that the ‘modules’ for processing music in each brain are slightly different, accounting for every person’s unique experience of music, and also for the varying levels of enjoyment individuals may derive from a performance (Chase 2006:17). Chase says that these ‘modules’ are developed as one grows and are influenced by environmental factors such as culture to determine your unique musical palette (2006:17). However, there do seem to be some universal musical traits that cross cultural boundaries and affect listeners regardless of the individual. Some of the universal/inherent qualities of music are: understanding of cadence, emotional affect, harmonic sensing (sensing the relationships between notes - intervals with small-integer ratios are favoured), melody, phrasing, rhythm, scales and repetition (2006:49). The more of these universals one can manipulate simultaneously in a performance, the greater the chance that an emotional response would be evoked.

Recent studies have shown that the brain processes music in the following sequence: First, the sound is received in the auditory cortex for initial processing. It is then divided into its component parts, and the frontal regions process the sound for musical structure and create expectations based on past experiences. It uses 'feature extraction' and 'feature integration' to process auditory stimuli. Then the mesolimbic system triggers arousal and pleasure through the release of opioids. The cerebellum is active throughout, mainly processing the sound for rhythmic content (Levitin 2006:103,191).

Another interesting consideration is that we now know that the brain is able to become better at musical processing – its structures can learn and self-modify in order to process more efficiently (Levitin 2006:108). This suggests that a music listener can develop a greater appreciation for musical styles other than those he/she is accustomed to experiencing through increased exposure and experience of these 'new' musical styles. Studies have also shown that children who take only a few music lessons possess more advanced neural circuitry than those who do not (Levitin 2006:194; Chase 2006:36).

It has also been discovered that the memory centre of the brain is active during the listening process. Memory is the brain relocating the pathways it created and used on first perception of a particular event. It is thought that the brain actually encodes relative features such as the tempo and melody of a song when we hear it for the first time, and, because of the close link between the memory and emotional systems, it then relates and links the musical events to the other things the listener was experiencing at the time. Hence, we sometimes are reminded of a particular food, occasion, or experience when we hear a particular tune.

Having said all of this, the affective nature of music must be a product of clever manipulation of these functions of the brain. If music is, as Levitin says, a 'perceptual illusion', a composer must use how the brain perceives musical sound to his advantage to create music that 'speaks' to the listener as the music the composer writes is really re-created in the brain of the listener. The primary way in which a composer may do this is through the creation of anticipation or expectation in the listener. As Meyer (1956:14) states:

> *Emotion or affect is created when a tendency to respond is arrested or inhibited.*

By this he refers to the ability of music to create or activate tendencies and expectations (based on the listener's previous musical experiences), inhibit those tendencies, and then ultimately provide meaningful resolution to them (Meyer 1956:23; Levitin 2006:65). This means that the listener must have some pre-existing idea of the musical conventions which accompany a certain style of music in order to be surprised, excited or inhibited by the musical material at hand (Deutsch 1982:318).

Deutsch (1982:402) shows that the mind is constantly working at predicting the possible progression and conclusions when processing music. With this idea in mind, the logical assumption must be that through greater building of this expectation or suspense in the listener, the composer can achieve a greater emotional release when the suspense is concluded and the mind of the listener reaches the state of completion, stability and rest that it so desires (Meyer 1956:128).

Xenakis (1971:76) offers advice for composers in terms of creating a greater affective response in the listener. He postulates that exciting the listener away from the presented energy holds musical interest.

With this information at hand then, one may ask why we do not grow tired of music we have heard many times over, as, if our expectations are what evokes an emotional effect then when we know what is to come we should not be affected by it anymore. Levitin (2006:105) answers this question by showing that the brain, even when it knows a piece of music well, still goes through the same processing system, and thus still creates the same expectations and suspense automatically. So, even when a listener knows what is about to happen he/she will still experience its affect.

This thought that musical affect comes from the listener's sense of expectation is not new. Langer (1942) states that musical affect is evoked through tension and release; motion and rest; agreement and disagreement; preparation; fulfillment; excitation; and sudden change.

Levitin (2006:172) links these devices to the common elements of music in saying:

Music communicates to us continually through systematic violations of expectations. These violations may occur in any domain – the domain of pitch, timbre, contour, rhythm, tempo, and so on – but occur they must.

So, music's affect is achieved by the contrast between the satisfaction of the expected coming to pass, and the tension created when the unexpected happens. An emotional response of some sort is the normal reaction to this experience.

The responses of test subjects to the listening materials in the experiments in this study will, according to the writers above, come from their experience of the material based on the individual expectations and predictions they generate during the test. These expectations are difficult to measure except through a style of questioning which is beyond the scope of this research however I feel based on the evidence above that expectations will still play a major role in dictating the responses to the listening material.

Chase (2006:172) shows that emotions have two main characteristics: valence and intensity. The 'valence' of an emotion refers to the type of emotion, positive or negative (emotions cannot be neutral), and then a more specific signifier such as 'sadness' or 'joy'. 'Intensity' of emotion means exactly that – the force of the particular emotion experienced. It asks: 'HOW much anger did you feel?' The answer may be anything from mildly angry to extremely angry. Valence and intensity (or arousal) are concepts that are also explored in the experiment design section of this work, as they become axes on which emotions can be measured using Thayer's model.

The phenomenon referred to as strong experiences with music (SEM) is experienced by many music-listeners and describes the power music can have in evoking a strong emotional reaction in its listeners. The mechanism of SEM is still unknown. In research designed to reveal this mechanism, it was found that listener-ratings of SEM were strongly related to their experiencing of five main physical reactions to musical material. These were: goose pimples (the feeling of hair standing on end), lump in the throat (difficulty in swallowing), shivers down the spine (or chills), being close to tears, and arousal (a state of physical anticipation). The result of the research indicated that SEM is strongly related to a synthesis of these five physical reactions at the conclusion of a musical piece (Yasuda 2008:1).

Numerous studies have investigated the acoustic features of music which cause physical reactions in listeners. In a survey on physical reactions evoked by music-listening, Sloboda (1991) found that many participants had experienced shivers down the spine, laughter, lump in the throat, and tears during listening. The results also suggest that some of the structural features of music induce tears and shivers down the spine. His research shows that tears are often caused by "melodic appoggiaturas" and "melodic or harmonic sequences". Particular acoustic features such as "new or unprepared harmony" and "sudden dynamic or textural change" frequently cause shivers down the spine.

Among the physical reactions to music that have been investigated, the phenomenon of chills (e.g. shivers down the spine and goose pimples) has been the focus of many studies trying to determine the acoustic features which induce them. Panksepp (1995) found that intense and dramatic crescendos caused chills. Grewe, Nagel, Kopiez, and Altenmüller (2005) concluded that harmonic sequences, the entrance of a voice, and the beginning of a new part of the piece which violated the expectations of the listener were the musical factors important for inducing chills. Nagel, Kopiez, Grewe and Altenmüller (2008) indicated that chills were frequently reported with a significant increase of loudness in particular frequency ranges. Grewe, Nagel, Kopiez, and Altenmüller (2007) reported that the entry of a voice and the volume (fortissimo, pianissimo) or a change in volume (crescendo, diminuendo) could be a trigger for chills. Guhn, Hamm and Zentner (2007) suggested that the following six musical and/or acoustic features can provoke chills:

- slow tempo
- alternation or contrast of a solo instrument and the orchestra
- sudden or gradual volume increase from piano to forte
- expansion in frequency range in the high or low register
- harmonically peculiar progressions that potentially elicit a certain ambiguity in the listener
- specific interplay of the melodic and harmonic progression.

Goose pimples, lump in the throat, shivers down the spine, and arousal were strongly related to the magnitude of change in volume. Therefore, these four physical reactions tended to be induced according to the degree of crescendo. Among these physical reactions, goose pimples and shivers down the spine were related to the magnitude of change in volume, which supports the findings of some previous studies that chills were aroused by the magnitude of change in volume (Panksepp, 1995; Nagel et al. 2008; Grewe et al. 2007; Guhn et al. 2007). In addition, they present results that reveal that lump in the throat and arousal were also closely linked to the magnitude of change in volume with crescendo.

What is problematic in all of these tests is that although the presence of the 'musical effect' has been confirmed by such experiences as 'shivers' and 'chills' and so on, it is still not clear as to whether or not the respondents understood what they were listening to in the terms used to describe the features of the music they were exposed to.

The following list of particular musical elements represents the observation that these seem to be the most thoroughly researched in music-emotion studies, and most frequently mentioned components of music by academic musicians. However, as alluded to earlier in this section, these may not be the elements of music that are actually perceived or understood by untrained listeners.

- Pitch
- Notes
- Melody
- Harmony
- Tone colour/Timbre
- Mode
- Scales
- Time
- Rhythm
- Tempo
- Form
- Contour
- Image Processes
- Articulation
- Loudness

- Dynamics
- Ornamentation

Due to the technological nature of this project, a new method of classification of musical elements must be created, as a computer does not understand such musical concepts as ‘note’, ‘key’, ‘scale’, and ‘rhythm’. For this reason I have divided the elements of music into 3 categories: Time-based elements, Amplitude-based elements, and Frequency-based elements. These are all simple terms (linked to easily measurable components of sound) in the world of digital computing, and I have found that almost all elements of music can be contained within one of these categories. I have also found that traditional musical elements that fall into the same category will often be inter-connected, thus rendering these categories even more helpful.

3.1.2 Frequency-based elements

Frequency-based elements of music are those components of music that are dependent on the presence of a particular frequency (or set of frequencies) for their successful transmission to the listener.

3.1.2.1 Pitch

Pitch is the psychological equivalent of frequency. Pitch exists only once one's brain has analyzed the speed of the vibration of a stimulus (Levitin 2006:22). In music, pitch signifies a steady state of vibration. To be perceived as “pitched”, vibration must be regular, and long enough to be recognized as regular by the human ear i.e. even some sounds we would consider noises might have a recognizable pitch, thus pitch is not an exclusively musical trait (Woods 1975:1). Higher pitches are more quickly identified as they contain a higher rate of cycles per second and so the brain may process them faster. Other factors such as amplitude, complexity of waveform, noise, and natural delays also have bearing on speed of perception of pitch (Woods 1975:44; Macone 1990:103). Roederer refers to pitch as being the perceived “height” of a sound (1979:4).

Due to pitch being the speed of vibration of an object, we could say that pitch is merely a much faster version of rhythm, as a rhythm is a constant vibration too (Woods 1975:42).

Tests show that high pitches are generally perceived as positive emotions and include descriptors like happy, graceful, serene, dreamy and exciting; and can also be associated with surprise, anger potency, fear and activity. They are also perceived as more intense than low sounds (Deutsch 1982:161; Chase 2006:172). Low pitches bring feelings of sadness, dignity/solemnity, vigour and excitement as well as boredom and pleasantness. Large variation in pitch in a piece of music shows happiness, pleasantness, activity or surprise. Small variation shows disgust, anger, fear, or boredom (Gabrielsson and Lindstrom 2001; Chase 2006:172).

Pitch should not be confused with notes, as although all twelve notes in Western music have pitches assigned to them, not all possible pitches (frequencies) have notes assigned to them. The distance from one note to the next in our Equal Temperament system is defined as 100 cents.

3.1.2.2 Notes

Notes are possibly the most rudimentary part of music. A note is a sound of defined pitch. An interesting fact regarding notes is observed by Chase (2006:177). He says that humans automatically and effortlessly use discrete pitches when speaking or singing, whilst our closest relatives, primates, vocalize in 'pitch glides' or grunts. These discrete pitches serve to allow the brain to organize sounds and recognize patterns in order to make sense of them.

In Western culture we use the "equal temperament" system, which affords us twelve different notes in the octave. Notes are used to create the melodic and harmonic content of music.

The power of notes is discussed by Copland (1963), who speaks of a single note having the ability to change the atmosphere in a silent room. The use of notes in music will be discussed in greater detail in the following sections, especially melody and harmony.

3.1.2.3 Melody

The melody of the musical composition is what the layman would call the "tune". Melodies are composed from sequences of notes moving in linear fashion (as opposed to the vertical stacking of notes in harmony). Copland notes that the melody of a song is perhaps the most superficial layer of music, or, the most easily noticeable to the untrained musician (1963:5). Roederer calls the melody in a piece of music the "musical message" which may or may not have meaning.

Macone states that melody is a "partial expression of tonality", that "a line of melody, whatever else it may signify, is a proposition concerning the interpretation of data in succession" (1990:103), and that it is "by nature, dynamic and transitional, driven by impulses and sensations which are not always subject to rational control" (1990:93).

In order for a melody to be successfully translated from the composer's pen to the listener's brain, it requires effort from composer, performer, and listener -

A listener's perception of melodic succession at any one moment will depend on how the composer and performer jointly determine the perceived quality of succession or simultaneity, and also how the individual listener manages the information (Macone 1990:103).

Macone goes on to say that the expressive implications of a melody are determined by the pattern of direction change, and the magnitude of that change. The possibilities for melodic material are limited by some physical issues, "it is restricted to the range of individually distinguishable pitches … and it is also under a speed restriction if each pitch is to be clearly identified" (Macone 1990:104). Thus notes available for use in a melody are limited to about the range of the piano keyboard, as these notes span the normal range of human discernment. Melody should also be played with some amount of feeling by the performer, as the absence of melodic inflection is construed by listeners as an absence of subjective interpretation. That is, listeners can perceive the difference between a 'musical performance' of a work, and someone merely playing through the notes. The listener is thought to require or desire some individualism in the performance for it to be considered effective (Macone 1990:106).

There are various documented techniques for creating a sense of expectation or suspense in the listener through melody. The initial idea is that the brain usually wishes to hear notes in a particular succession (which it judges on past similar experiences and learning), and makes judgments as to what note should follow the previous one. By simply omitting a note from this 'good' or expected succession a measure of suspense is evoked in that the listener's desire to hear the note that was omitted is not satisfied. The composer can then allow the desired note to be heard at a delayed stage, or to take the melody even further away from that note to build a greater desire for it (Meyer 1956:131; Levitin 2006:118). Melodies that use notes that are not from the tonal center or 'key' of the composition (called 'chromatic notes') are particularly effective in achieving this effect.

Gabrielson and Lindstrom also found that melody contributes to the emotional effect of music. They cite studies wherein various qualities of melodies were tested, and their effects noted by the test subjects.

Melodic range:

A wide melodic range was mostly associated with feelings of joy, whimsicality, and uneasiness, whilst a narrow melodic range was seen to evoke feelings like sadness, dignity, sentimentality, tranquility, delicacy and triumph.

Melodic direction:

Melodies of ascending nature were felt as dignity, serenity, tension and happiness (but not in test conducted with children) and fear, surprise, anger and potency. Descending melodies were exciting, graceful, vigorous and sad (especially when in a minor mode) and also evoked boredom and pleasantness.

Melodic motion:

The interaction between contour and rhythm affected judgments of happiness and sadness (Gabrielsson and Lindstrom 2001).

3.1.2.4 Harmony

In music, harmony refers to the vertical combination of notes, or "the combination of two, three or more voices or instruments sounding at different pitches" (Macone 1990:167). There are two main chord families: major and minor, each with many possible variations.

Harmony presents an interesting dilemma, as harmony is the combination of single notes, thus the listener can hear both the single notes, and the 'fusion' of these which usually possesses a different quality than any of the single notes. There is little focus in musical analysis on the fusion of notes and the affect it evokes, as most of the attention in this area is given to the discrete components involved in these fusions (Smith 1979:241).

Copland (1963) notes that harmonic movement in music generally escapes the notice of the average listener, but does this mean it has no effect on this listener? If harmony is seen as the movement of separate horizontal lines of notes moving simultaneously, on two or more vertical planes, then we could say that the average listener must at least perceive the harmony's interaction with the melody (this is, in any case, how harmony is analyzed), even though the melody is most important in his/her mind.

Although there is currently no absolute agreement as to why some intervals are considered consonant and others dissonant, Roederer (1979:5) and others believe that the perception of harmony is dependent on cultural conditioning. In general, for Westerners, harmonic consonance has a pleasing quality, whilst harmonic dissonance is irritating and creates the urge to resolve to consonance.

Meyer (1956:182) says that sections of music with increased harmonic movement help to evoke a sense of suspense and expectancy in the listener and Macone (1990:174) attributes the following effects to harmony: “identity, stability, permanence, mutual co-operation, proportion, influence, efficiency, power.”

Gabrielsson and Lindstrom (2001) shows the results of tests to determine the perception of various harmonic contents. They found that simple, consonant harmony is perceived as happy, relaxed, graceful, serene, dreamy, dignified, and majestic; and that complex and dissonant harmony evokes excitement, tension, vigour, anger, sadness, and unpleasantness.

Diminished and augmented harmonies are thought to be heard as the least stable due to the difficulty in perceiving the root of the chord (Meyer 1956:226).

3.1.2.5 Tone colour/Timbre

The timbre, or tone colour of an instrument is dictated by high-order harmonics generated by the instrument’s design, and the playing technique employed to produce sound, and also by the mechanical noises generated in playing the instrument (e.g. the sound of the hammers on piano strings or finger-noise from a guitarist’s hands on the strings). Our ears recognize the sounds of different instruments through processing this harmonic information and associating that data with past experiences of various sounds to find a match (Roederer 1979:5; Macone 1990:36).

Copland (1963:23,24) and Levitin (2006:54) show that recognition of an instrument’s timbre may be a direct factor in creating an emotional effect as the sound of a particular instrument may hold a specific emotion for some people. Copland also mentions that tone-colour can be bigger than the sound of a single instrument, and that in ensemble settings, timbres may be mixed by a composer, giving birth to many new possibilities. In this instance the ensemble may be seen as one instrument being played by the composer. Deutsch (1982:12,14,42) also mentions that specific timbres may affect changes in perceived meaning of a musical phrase or performance, and also that the context in which a particular timbre is placed affects how that timbre is perceived.

Gabrielsson and Lindstrom (2001) cites tests which showed the effects of tone colour on listeners. The tests found that instruments with many harmonics incite potency, anger, disgust, fear, activity or surprise; amplified higher harmonics suggest anger. Few, low harmonics evoke sensations of pleasantness, boredom, happiness or sadness; suppressed higher harmonics bring tenderness or sadness. One study found sadness best expressed through singing voice or violin; the timpani was found to best produce associations with anger; and the violin was the best at creating a fearful sound.

Chase (2006:174) reports that instruments of simple tone colour (few overtones), like the flute, evoke pleasantness, peace and boredom; whereas a complex tone colour (many overtones) like an overdriven electric guitar is associated with power, anger and fear. A bright tone colour is associated with positive emotions, mostly happiness, whilst the opposite is true of a dull tone colour. Similarly to Gabrielsson and Lindstrom, he also shows that the violin is best for evoking fear, as well as sadness and anger, and the drums (including timpani) are a good choice for evoking an angry response.

3.1.2.6 Mode

The “mode” of music refers to whether the music is in a major or minor key.

Meyer (1956:225) asserts that, due to its inherently greater potential for chromaticism, the minor mode has stronger tendencies to evoke an effect than the major, as less probability of harmonic progression can be expected by the listener – so the feeling of suspense is greater.

Gabrielsson and Lindstrom (2001) list mode as one of the factors that influences emotional response in a listener. In tests performed, it was found that, in general, major mode music was perceived as happy and minor as sad (even in 7 – 8 year olds). Major is also associated with feelings such as graceful, serene and solemn; and minor with dreamy, dignified, disgust, tension and anger.

Musical context was found to determine which of these traits was felt, for example, if high-pitched, loud chords were heard, more happiness was perceived than in low-pitched, soft chords, even in a major mode. This shows that musical elements must be co-dependent in order to evoke the desired response in the listener.

These studies must be placed in context, however, and I do note that while their findings hold true for the majority of western people, these are not universal truths (also in Levitin 2006:38).

3.1.2.7 Scales

Mode is closely linked to scale, as in Western music, when we employ the major mode for a work we also employ the major scale. The difference is that, where there are only two possible ‘modes’ for a piece of music (major and minor), there are many scales (see definition below) that fit into each of these modes. For example, by simply rearranging the notes of the major scale to start and end at different points, we can construct 6 new scales, all of which still fit the major mode.

Roederer (1979:161) says a scale is a “discrete tone sequence”, and that the particular scales used by composers and performers are mostly culturally based, as different cultures have grown to adopt different sounds for different meanings.

According to Chase (2006:239) a successful scale needs two elements: easily-processed simple-frequency ratio intervals (to provide a sense of stability and structure within the piece), and a few unstable, unbalanced intervals, especially a leading-tone (these are used to evoke tension).

Interestingly, almost all cultures have scales in their music and these scales have certain overarching organizing principles. Chase (2006:178) postulates that this is because when music initially developed and people began to compose, they automatically used notes that came ‘naturally’ to them. The tunes that were remembered best were then found to be the ones that used the same sets of notes. These sets came to be the scales of the culture they were originally used in.

The ancient Greeks developed a system of seven “modes” or scales, with each said to have a particular expressive trait. So, when a composer wanted to evoke a particular emotion, he would compose from the relevant mode (Macone 1990:95). This is a very direct way of evoking emotional response, the only possible shortcoming being that if your audience knows the mode, and the expected emotional response, does that allow them to truly feel what the music evokes, or only the emotion targeted by the composer? This leads one to the assumption that a certain level of integrity and discernment is required from a listener in allowing himself/herself to truly feel the music’s effect without a prior analysis of his/her expected response. Woods (1975:178) has also said that these emotions ascribed to the modes are probably not valid today because of our equal temperament tuning system. He does, however, postulate that certain keys may sound or 'feel' different on a particular instrument, as the mechanism whereby the instrument produces the various notes contained in the keys may differ.

Studies have also shown that scales must have an internal organization of small-integer or simple frequency ratios. Infants respond only to scales with this characteristic. This suggests that the human brain has a hard-wired preference for simple frequency ratios and that the construction of scales by this principle is not a cultural construct (Chase 2006:188).

3.1.3 Time-based elements

A time-based element is any component of music that relies on the constant passage of time in order to produce its effect.

3.1.3.1 Time

Macone (1990:66) speaks of music as ‘an experience of transience’ and says that ‘if sound did not die away, we would not have the option of repeating, or better still, replacing it with something different. Without silence there would be no music.’ He speaks of ‘the drama of music’ being ‘played out in real time … events which live on the leading edge of a perpetual process of transition from the future to the past, fade and die at the very moment an audience is listening’.

The decay of sound is dependent on the passage of time. Without the decay of one note or phrase to silence, the next could never start. Roederer (1979:5) points out that steady, unending sound is annoying, and eventually it would fade into the background and we would not notice it anymore. Musical sounds change with the passage of time, thus to be musical, sound must have a time-dependency.

Although time is a natural phenomenon, it must be considered an element of music, as, without it music could not exist (Levitin 2006:125). Smith (1975:16) calls music not 'static', but 'ek-static' (or 'continually becoming'), and nothing can continue without the passage of time. Musicians need time, not only to sound notes, but also, in improvisational settings, to think of what the next note will be. This use of time may not directly impact on emotion, but time allows the creator of the music to conceive of the music that will evoke an emotional response.

We must also be careful that the concept of musical time is not confused with physical time, though they are related. Musical time is merely a measured abstraction of physical time, related to physical time by tempo. Thus time values in music cannot be measured in physical time without a specified tempo (Mazzola 2002:1028).

3.1.3.2 Rhythm

The beat of a song is essentially its pulse. Rhythm refers to how that beat is manipulated by the composer, usually creating repetitive patterns of time-values. Copland (1963:5,22) places rhythm as the second most obvious musical element to the “man in the street”. He also states that good musical flow is wholly dependent on rhythm.

In Deutsch (1982:171) we learn that predictability of rhythm is one of the important features of stability in music, and Meyer (1956:121,122) says that rhythmic syncopation (the use of unpredicted rhythms) may be used to evoke a feeling of urgency in music.

Gabrielsson and Lindstrom (2001) show that in tests performed on listening subjects using various rhythms at varied tempos, faster rhythms were generally seen as happier, and slower rhythms as sadder, irrespective the of particular rhythm. They also found that smoother, more regular rhythms evoked happiness, dignity, majesty, and peace, whilst irregular rhythms were perceived as amusing, uneasy, and angry. Varied rhythms evoked the feeling of joy. Firm rhythms sadness, dignity, and vigour; and flowing rhythms were happy, graceful, dreamy, and serene.

3.1.3.3 Tempo

The tempo of a piece of music is how fast or slow it is. Tempo is claimed to be the most influential factor in music that aids in evoking emotional response (Gabrielsson and Lindstrom 2001; Meyer 1956:15).

Studies of the effects of tempo on listeners found that slower tempos were perceived as more serious, and faster tempos as happier. More specifically, the faster the tempo, the higher the perception of activity/excitement, happiness/joy/pleasantness, potency, surprise, anger and fear. Slower tempos gave rise to feelings of calmness/serenity, dignity/solemnity, sadness, tenderness, boredom and disgust (Gabrielsson and Lindstrom 2001:235). Other factors such as the mode (major/minor) may override this general trend though (Gabrielsson and Lindstrom 2001:236).

It must also be said that "subjective tempo" played a part in the perceptions, as denser music was generally perceived as being faster than music with a sparser sound-palette (Gabrielsson and Lindstrom 2001).

3.1.3.4 Form

Form is the arrangement of bars of music into a structured composition. Although it is an essential element of music (Copland 1963:5), studies in which the effect of form was tested through changing the order of movements of popular classical pieces found it had little effect on perception or emotional response in listeners. It would seem that the listener does not take much notice of the overall structure (probably as it is simply too large), but rather focuses on the more immediate elements being presented (Gabrielsson and Lindstrom 2001). Meyer (1956:134) does however reference the power to evoke affect by creating a 'structural gap' and not filling space in the form as the listener would expect.

3.1.3.5 Contour

The term “contour” in music refers to the degree of vertical change in a line of music (usually the melody). Macone (1990:70) states that the melodic contour of a piece may affect perception of tempo, with more dramatic changes in contour being perceived as faster than more gradual changes.

The evidence that we are constantly analyzing the contour of a melody is shown in Deutsch (1982:415), in that even infants responded to a change in the contour of a melody they were accustomed to, but did not respond differently when the whole melody was simply transposed to a different key.

Contour is a time-based element because the vertical change in the musical line occurs over a period of time.

Meyer (1956:160) shows that where good contour is lacking in music the listener will have the urge to improve it. This is an urge he/she cannot satisfy and is thus a strong means of arousing tension.

3.1.3.6 Image processes

This is a time dependent element in a different manner. Meyer (1956:10) postulates that the meaning we find in music may arise from an association made in our minds to something we are reminded of from previous experiences, thus the passage of time is required to recall past experience.

He calls this method of experiencing music an “image process”. This would dictate that the content of a listener’s mind at the time of listening may affect his/her perception of the material. It also indicates that a given piece of music could mean a different thing to every listener in an audience as the meaning is generated in each mind from its own experience. He does, however, state that certain musical traits such as culturally related images can easily relate to a large group of similar culture, and that certain feelings or images can be evoked through idiomatic clichés (like “Africa” is associated with tribal drum music). He notes that various instruments have come to have images linked to them and that specific tunes may hold a specific meaning for individuals or groups (Meyer 1956:10,12; Macone 1990:43). He also asserts that the goal of music may not be to evoke a concrete image, that it need only evoke a mood/feeling to be a success.

Another way in which the brain creates an image is when we listen to recorded music and we are often able to construct a likely image of the space in which the music was recorded, due to the brain's ability to use reverberation cues to create a spatial image, and its use of frequency analysis to imagine the structural materials of the venue.

3.1.3.7 Articulation

Articulation is how long or short a performer holds a note for. For example, a ‘staccato’ (short) crotchet – supposed to last one beat in normal time – may be played for only a fraction of that beat to impart a different feeling to the note. Likewise, a ‘legato’ (long) crotchet may be held a bit longer than its full value.

Articulation is the largest component of 'phrasing' in music, and musicians often interpret musical material in an individual way to bring freshness to the content. In studies of the phrasing of various singers (Seashore 1938 in Deutsch 1982) it was noted that phrasing was altered (through changes in the length of syllables) instinctively by the singers when they were asked to evoke particular emotions (Deutsch 1982:93).

Staccato notes may be associated with “gaiety, energy, activity, fear and anger”. Legato may evoke the sensations of “sadness, tenderness, solemnity and softness” (Gabrielsson and Lindstrom 2001).

Another possible meaning for articulation may be 'the physical movements the performer makes as he/she performs'. Studies have shown that audiences do feel that they receive emotional cues from the performer's movements, and that performers also believe that their movements can affect the perceived meaning of the music they play (Levitin 2006:210; Nattiez 1990:44).

3.1.4 Amplitude-based elements

An amplitude-based element is any element in which a variation in amplitude is the stimulus for the element’s effect.

3.1.4.1 Loudness

Loudness is the perceived strength, intensity and volume of music (Macone 1990:81). A variation in loudness is achieved through alteration of the power of a sound wave.

Gabrielsson and Lindstrom (2001) show findings of tests on the perceived effects of loudness. “Louder” music was perceived as having greater intensity/power, tension, anger and joy, and “softer” music had the feeling of softness, tenderness, sadness, solemnity and fear. Large variations in loudness may suggest fear, and small variations happiness or activity. Rapid changes are playful, or pleading. Few or no changes throughout a piece were associated with sadness, peace, and dignity. Chase (2006:173) includes that extremely loud music in the range of distortion is perceived as angry, and music with sudden changes from soft to loud evoke fear.

3.1.4.2 Dynamics

Dynamics is the flow of changes in volume and intensity (indicated on the score by the composer, and interpreted by performers) of a piece of music. Meyer (1956:15) says that the dynamic content of a piece is critical to evoking imagery in the listener. Macone (1990:43) mentions the use of dynamics as important for composers in grabbing the attention of the listener, and then using dynamic changes to shift his attention to an intended point. Thus we see that a shift in amplitude may have a direct effect on what the listener perceives, as when one's focus is shifted, one also hears differently. Composers may use this to increase the likelihood of the listener feeling the intended emotional effect by directing attention to the particular place where that effect is most dramatically created.

3.1.4.3 Ornamentation

"Ornamentation is the convenient term found in dictionaries of music to describe various ways in which a performer brings freshness, spontaneity, and a sense of individual personality to a familiar composition" (Macone 1990:122). This means that ornamentation is the way in which a performer embellishes a written musical line through various techniques, to add colour and interpretation to the music. Macone (1990:122,123) talks of ornamentation being used by performers to evoke surprise and excitement, dynamism, and flow – certainly feelings that are linked to the creation of an emotional response to music. Meyer (1956:206) says that ornaments are used to evoke psychological tension. Ornaments are often combinations of the other elements discussed above, for example, the vibrato effect is a combination of the variation of pitch and loudness (Woods 1975:58).

Although this list of musical elements and the discussion thereof provides some insight into the components of music that aid in creating the 'musical effect', a growing body of research has been emerging wherein the goal is to identify the emotion of a piece of music by digital analysis of its sonic features, or to synthesize emotion through understanding the more global parts of the sonic whole which are common to music and are believed to have a particular emotional effect. Due to the fact that this study has goals similar and related to these – but with noise as the agent to be manipulated – these fields of research must be investigated and explored. Points of correlation with the elements of music already discussed will be identified, and, once this is achieved, some insight into the elements which could be created via digital techniques should begin to surface.

3.2 MUSIC EMOTION RECOGNITION (MER)

Over the last two decades or so researchers have given increasing amounts of attention to attempting to uncover methods whereby music can be analysed (by digital means) and the emotional content of the music detected (Panda and Paiva 2011). Especially since 2007 this work has garnered much attention as the Audio Mood Classification (AMC) task has been run annually at the Music Information Retrieval Evaluation eXchange (MIREX), the community-based framework for the evaluation of Music Information Retrieval techniques. The field dealing with the prediction of emotion from a music signal is usually called 'Music Emotion Recognition'. The primary use of this prediction and analysis is to develop methods for categorising music libraries by emotion rather than by artist. This would serve as a tool for listeners to select music-listening based on their current mood – or the mood they desire to be in. MacDorman (2007) believes that the ability to select music in this way would allow listeners to appreciate their music collections better.

Since the advent of the personal MP3 player and the proliferation and availability of music online more and more people have made the move toward keeping their music library on either a computer or mobile device such as an iPod. People working in the MER field saw this as an opportunity to catalogue music collections in new ways, as the music now exists in the same domain as the tools for processing and analysis. The tradition of CD's holding one's music limited cataloguing-potential due to the 'hard-copy' format of a CD.

Other potential uses for this type of music analysis are: music for film could be guaranteed to evoke the emotion that is intended; background music in restaurants and shopping malls could be chosen more appropriately; and music could be used more effectively in education (MacDorman 2007).

I believe this field could be of use, as the initial objective of this field was (and still is) to ascertain what features of the music could be analysed in order to determine its emotional content. It stands to reason then, that if musical features can be analysed and the emotional response to music predicted and the predictions prove accurate, then one could create music with predetermined features that have been shown to evoke certain emotions – a sort of 'reverse-engineering' of the process in order to fulfil this project's goal of manipulating noise to evoke the 'musical effect' which Han (2009) classifies as ‘transition of emotion state by music’.

Thus, to follow is an investigation into the field of MER with a focus on its attempts to show which features of music evoke which emotions in listeners, how these features are analysed, and how the emotion-quotient of music is measured.

MER projects are generally undertaken through the processing and analysis of large batches – into the thousands – of short segments of music at a time. The music is all converted to segments of equal length and sampled at the same rate to ensure consistency across the audio signals. Computers run algorithms – usually from programs such as PsySound or MARSYAS (developed to classify musical genre) or MIRToolBox (sometimes multiple programs are employed simultaneously) – which measure multiple values for the various acoustic features of the music which have been deemed good measures of emotional information or have a correlation with musical features which are good measures of emotion. The computers' predictions of emotion are then tested on human subjects in order to confirm or refute the predictions and thereby refine the algorithms and processes used (MacDorman 2007; Pohle 2005; Wu 2006; Trohidis 2011; Trohidis 2008; Yang et al. 2007; Kim 2010; Yang and Lin 2008; van de Laar 2006).

Acoustic and musical features which have been found useful for emotion-mapping are:

- tonality (MacDorman 2007)
- stability (MacDorman 2007)
- intensity (MacDorman 2007; Trohidis 2011)
- perceived pitch height (MacDorman 2007)
- loudness (MacDorman 2007; Kim 2010; van de Laar 2006)
- change in pitch (MacDorman 2007; van de Laar 2006)
- tempo (Paleari 2007; Kim 2010; Han 2009; van de Laar 2006)
- harmony (Kim 2010)
- roughness (Kim 2010)
- notes per second (Kim 2010)
- attack time (Kim 2010)

- peak sound level (Kim 2010)
- articulation (legato/staccato) (Han 2009; van de Laar 2006)
- dynamics: overall level, crescendo/decrescendo, accents (van de Laar 2006)
- timbre: spectral richness, onset velocity, harmonic richness (van de Laar 2006; MacDorman 2007; Trohidis 2011; Kim 2010)
- interval (small/large) (van de Laar 2006)
- melody: range (small/large), direction (up/down) (van de Laar 2006; Paleari 2007)
- harmony (consonant/complex-dissonant) (van de Laar 2006)
- tonality (chromatic-atonal/key-oriented) (van de Laar 2006)
- rhythm (regular-smooth/firm/flowing-fluent/irregular-rough) (van de Laar 2006; MacDorman 2007; Trohidis 2011; Paleari 2007)
- mode (major/minor) (van de Laar 2006; Paleari 2007)
- musical form (complexity, repetition, new ideas, disruption) (van de Laar 2006)

The main problem with using some of these features is that the technology is not yet able to accurately and consistently detect some of the higher-level attributes, such as harmony, melody and tonality (Yang and Lin 2008). In most cases the solution to this seeming problem is to use the 'lower-level' acoustic features that can be calculated from the raw audio waveform of a musical piece. In this way, an algorithm extracts numbers and figures out of the music. Different categories of these features have emerged, and differ from low to high level and complexity (van de Laar 2006).

The acoustic features tested by many of the researchers in the MER field are largely the same set, as these are the ones they have been able to test through the use of the available technologies.

Kim (2010) arranges musical features and their properties so as to understand which musical features may be discernible from analysis of acoustic features of a music signal, in a table called 'Common acoustic feature types for emotion classification'.

Type	**Features/Characteristics**

Dynamics	RMS energy
Timbre	MFCCs, spectral shape, spectral contrast
Harmony	Roughness, harmonic change, key clarity, majorness
Register	Chromagram, chroma centroid and deviation
Rhythm	Rhythm strength, regularity, tempo, beat histograms
Articulation	Event density, attack slope, attack time

Hu and Downie (2010) chose to use the audio features selected by the MARSYAS submission to MIREX because it was the leading audio-based classification system evaluated in both the 2007 and 2008 Audio Mood Classification (AMC) task. MARSYAS used 63 spectral features including means and variances of Spectral Centroid, Rolloff, Flux, Mel-Frequency Cepstral Coefficients (MFCC). Although there are audio features beyond spectral ones, spectral features were found to be the most useful and most commonly adopted for music mood classification because the classifier operates on a time-based system. Spectral analysis is performed in time parcels and then sent to the classifier for analysis.

MacDorman (2007) takes a similar approach to feature extraction, focussing on three features of audio he felt characterized a genre: timbre, pitch, and rhythm. Mel frequency cepstral coefficients (MFCC), the spectral centroid, and other features computed from the short time Fourier transform (STFT) were used in the extraction of timbral textures. A beat histogram represents the rhythmic structure, while a separate generalized autocorrelation of the low and high channel frequencies is used to estimate pitch. Three of the methods – the spectrum histogram, periodicity histogram, and fluctuation pattern – are derived from the sonogram, which models characteristics of the outer, middle, and inner ear.

Han (2009) shows a method that may allow chords and scales to be analyzed using chromagrams and spectrograms. A chromagram of each frame indicates where the notes are concentrated. Based on this, chords can be computed using key profile matrix. Similarly, scales can be computed by grouping and comparing intensities of scales. After detecting chords and scales their differences are computed for the current and next frames. This is a promising initiative, however it seems to be in its infancy and thus not yet totally reliable.

Van de Laar (2006) gives simple explanations of the various acoustic measurements:

Firstly there are the musical surface or timbral texture features. Most of these features are based on a Short Time Fourier Transformation (STFT).

Centroid

This is the mean of the Short Time Fourier amplitude spectrum. It gives an indication of how bright a musical piece is.

Roll off

This is the point where frequencies are getting smaller in amplitude and gives the shape of the spectrum. 95% of the total spectrum is within this range.

Spectral Flux

This indicates how much the spectral shape changes from frame to frame.

Zero Crossings

This feature gives the number of times the signal crosses the zero line. It is a good indicator of the amount of noise in the signal, and also of the fundamental frequency of the signal.

Low Energy or Average Silence Ratio

This represents the percentage of frames with less than the average energy of the entire signal.

Spectral Flatness Measure

This acoustic feature quantifies how tone-like a sound is. It is based on the resonant structure and the spiky nature of a tone which is quite different compared to the flat spectrum of a noise-like signal.

Spectral Crest Factor

This is the ratio between the highest peaks and the mean RMS value of the signal. This feature can be used in different frequency bands and quantifies how “spiky” the signal is.

Mel-Frequency Cepstral Coefficients

Mel-frequency cepstral coefficients (MFCCs) are coefficients that are collectively called a Mel-Frequency Cepstrum (MFC). They are derived from a cepstral representation of an audio excerpt - which is a nonlinear "spectrum-of-a-spectrum". The primary difference between the standard cepstrum and the mel-frequency cepstrum is that in the MFC, the frequency bands are equally spaced along the mel scale, which resembles the human auditory system's response more closely than the log-spaced frequency bands normally used. This can allow for better representation of sound.

MFCCs are derived as follows: The Fourier transform of a signal is taken and its spectral powers mapped on the mel scale using overlapping windows. Then the log-value of the power at each mel-frequency is noted. A discrete cosine transform of the list of mel log powers is made, as if it were a normal signal. The MFCCs are the amplitudes of the resulting spectrum (Min Xu et al. 2004; Sahidullah and Saha 2012).

DWCH

DWCH stands for Daubechies Wavelet Coefficient Histogram. This is a higher level and more complex feature than the above. Wavelets are a sort of histogram for sound. They divide up data, functions, or operators into different frequency components. There are various advantages to wavelets. Good time resolution at high frequencies and good frequency resolution at low frequencies - every octave lower down the scale normally has a lower resolution in terms of Hertz. Accuracy is improved as noise has less influence on the feature and there is less correlation between features so that a certain feature is less ambiguous in evoking an emotion.

These methods overlap significantly and have been widely adopted as 'the' measurements to be taken when attempting to predict emotion in music. One stumbling block here is that the best accuracy rate achieved for prediction of emotion in a music signal is around 65% (Panda and Paiva 2011), and this was achieved when using a combination of all features for prediction, not isolated features (Kim 2010). So any results in the data in the current research which approach or exceed 65% they could be seen as quite compelling, as this research will not test combined musical elements but individual ones which have been even less successful in achieving success in other studies. Van de Laar (2006) concludes that there will probably never be a perfect system for doing this work, but that the methods presented above do suffice. Hu and Downie (2010) investigated using lyrical/textual content as additional information to be used when predicting the emotional content of a musical piece. Although it does seem to improve the results of prediction somewhat, it is unsuitable for the purpose of this project as lyrics are not part of 'noise'. Thus, it is clear that though some promising results have been yielded, technology is not yet able to reach a very good rate of prediction of mood in a music signal. Interestingly, in one area of Thayer's Arousal-Valence emotion-model, Hu and Downie (2010) found that lyric-based features could not improve on the results obtained by using purely acoustic features for emotion prediction. This was the negative valence, negative arousal (represents emotions such as 'calm') quadrant of the model.

One other problem that researchers have experienced is that the socio-cultural context of the listener has a definite role in his/her perception of emotion in music (Pohle 2005). While this may be a problem in identifying specific emotions in a musical piece it does not affect the proposed idea behind this research, as no matter the socio-cultural background of the listeners, they should be able to identify whether a noise-signal moves them emotionally or not, or whether the emotion in the signal shifts over time or not.

3.2.1 Emotion-Mapping Models Commonly Used in MER Listening Tests

Listening tests involved with identifying emotions generated by music have generally been administered by asking listeners to ‘rate their emotional response using a validated index, that is, one with high internal validity. It is worthwhile for us to construct a valid and reliable index, despite the effort, because of the ease of administering it' (MacDorman 2007). This ease of administration to which he refers is in comparison with the method of measuring respondents' vital statistics electronically throughout the experiment in order to monitor emotional changes – a method which requires the use of expensive technologies which often require recalibration for individual users making for an arduous and expensive process.

The study of emotion in music began in the late 19th century but has only been pursued seriously from the 1930s. The results from many studies demonstrated strong agreement among listeners in defining basic emotions in musical selections, but greater difficulty in agreeing on subtle emotional nuance. Personal bias, past experience, culture, age, and gender all play a role in how an individual feels about a piece of music, making classification difficult. Due to the wide acceptance that music expresses emotion, some studies have proposed methods of automatically grouping music by mood (MacDorman 2007; Yang et al. 2007; Trohidis 2011; Panda and Paiva 2011; Pohle 2005; Wu 2006; Trohidis 2008; Kim 2010; Yang and Lin 2008).

Studies on emotion in music revealed a strong agreement on the effect of music on two fundamental dimensions of emotion: valence (or pleasure) and arousal. The studies also found agreement among listeners regarding the ability of pleasure and arousal to describe accurately the broad emotional categories expressed in music (MacDorman 2007). These findings have led many scholars in the MER field to adopt a dimensional approach to mood categorisation based on Thayer's 1989 interpretation of Russell's (1980) 'Circumplex Model of Affect' which uses valence and arousal as the two axes against which mood or emotion is plotted. Valence can also be called pleasure or evaluation and speaks to the positive/negative association in the music, and arousal is also called activation and is a measure of calm versus excitement (Wu 2006). This two-dimensional model results in a graphic representation with four quadrants: high valence – high arousal; high valence – low arousal; low valence – low arousal; and low valence – low arousal. Upon listening to a piece of music respondents are asked to plot the valence-arousal position in terms of which quadrant it is in, and sometimes also its position in the quadrant, relative to the intersection of the axes, in order to indicate the perceived intensity of the emotion.

Yang et al. (2007) makes some important points regarding the listening tests, the first of which is that subjects are asked to label the emotion based on their feelings of what the music sample is trying to evoke, rather than the emotion the subjects perceive at the listening test (Trohidis 2011; Kim 2010). We must make this distinction clear because perceived emotion and evoking emotion are not always the same. For example, a person who enjoys music with a sorrowful tone might feel happy when listening to songs that exhibit this tone – however the song may not be evocative of happiness. If the subject responded that the emotion of the song was 'happy' because the tone makes him happy the data would be skewed. Since MER is being developed to aid retrieval of music through identification of the emotion they wish to hear/feel, it is more natural that the Valence-Arousal values of a song are correspondent with the evoking emotion. No limitations on background, culture or expertise are imposed when recruiting subjects since the MER system is expected to be applicable to common people. However, because music emotion recognition is still new to all subjects, they must be informed of the essence of the emotion model, the goal of the experiment, and the following rules of the subjective test:

1. Label the evoking emotion rather than the perceived one.

2. Express the general feelings in response to melody, lyrics, and singing (vocal) of the song. We do not attempt to ignore the influences of the lyrics and singing even though the related features are not considered so far – as stated above, lyrical content is not applicable to this work.

3. No limitation is given to the total duration of the labelling process - subjects are allowed to listen to the music samples more than once to ensure the labels can truly reflect their feelings – although typically the total duration of the labeling process is less than 15 minutes.

It has been found that, in general, the valence of a piece of music is more difficult to predict than its arousal (Tolos et al. 2005). Yang et al. (2008) gives two reasons for this. First, while there are a number of features related to arousal like loudness/dynamics, tempo, and pitch, few features have been found to indicate solely for valence. Second, individual perceptual difference for valence is larger than for arousal. There is a chance that two respondents could perceive opposite valence toward the same song (Yang et al. 2008).

Results of MacDorman's (2007) work showed that mid-range rhythms and medium-to-loud mid-range pitches tend to be much more pleasurable than low pitches and soft high pitches. They also showed that faster rhythms and louder higher pitches tend to be more arousing than slower rhythms and softer lower pitches. Wu (2006) discusses what he refers to as “Musician's Rules of Thumb” for the relationships between certain musical features and emotion. He says that a high arousal often corresponds to high loudness and a low arousal to low loudness. He also reveals that low loudness corresponds to quadrant 4 (emotions similar to "relaxed"). Broadly speaking, the upper half-plane corresponds to large loudness, which means that the active emotions (excited and angry, etc.) often come with music of large loudness. His general rule for tonality shows that only quadrant 2 is atonal (i.e., not corresponding to any of the 24 music keys, such as C major, A minor, etc.). The musician’s general rule for timbre says that bright sound is related to large arousal or more positive valence. Hu and Downie's (2010) work shows that quadrants 1 and 2 (high arousal) correspond to larger values, which means that the music there is brighter. Also, the third and fourth quadrants are related to lower value, indicating that music there sounds gloomier and softer. They also mention that timbre is also affected by the variation of the sound harmonic structure. Thus, he introduces a musician’s general rule for harmony, which has two poles: complex (dissonant) and constant (simple). Constant harmony may induce positive valence, and negative valence music mostly sounds complex and dissonant. A harmonic musical segment is often one with a low spectral dissonance. This means that negative valence sounds should be inharmonious. In their data only quadrant 2 is significantly related to inharmonic sounds. A musician’s general rule for tempo was also generated by this work, though it may be the most difficult to prove. The research found that high arousal emotion is generally expressed by fast music, and low arousal emotion is generally expressed by slow music. However, tempo is often perceived relatively and temporally. For example, if a former segment is relatively faster than a latter one, a listener may perceive the latter to be even slower than it really is. This makes it difficult to compare the tempos of different musical segments one by one in an experiment. Another problem is that tempo values are calculated by the beat histogram, which usually reflects a value calculated over the whole sound. Shorter musical segments may not be long enough to gather the strong peaks needed to make inferences as to true tempo. In addition, music from different genres often shows dissimilar rhythmic structure. For example, rock and hip-hop have clear dominant peaks in the histogram, while peaks in classical and jazz may be more obscure and the first or second harmonics could be mistakenly treated as the first peak. Hu and Downie's proposed solution to these problems is to collect samples from more types of music and build specific models for different genres.

3.3 MUSIC EMOTION SYNTHESIS (MES)

Music Emotion Synthesis (MES) is a field of research that deals with experiments in tailoring music to generate specific emotions in listeners, to generate representative music based on given emotional parameters and to invent new methods (instruments) that enable this process. Knapp and Cook (2005) argue that the primary goal of the music-performer is to express his/her thoughts and emotions through sound via interaction with a musical instrument and have this sound evoke an emotional response in the listener. Thus they define musical performance as the communication of emotion through sound. It would then follow that one definition of a musical instrument is a device that enables the expression of emotion through sound. MES seeks to understand how sound operates in the realm of emotion in order to facilitate the creation of emotionally evocative sounds. The major problem that the field must deal with is that, as Gunes et al. (2011) points out, despite major advances within the affective (emotion-based) computing research field in recent years, the digital modelling, analysis, interpretation and response to natural human affective behaviour still remains a challenge. He cites as a reason for this, the idea that emotions are complex constructs and are not necessarily clearly defined as single entities. Emotions also exist with substantial individual variations in expression and experience, further adding to problems of definition. A small number of discrete, simple categories (e.g. happiness and sadness) may not reflect the necessary subtlety and complexity of the various emotions conveyed by a rich source of information like music.

Knapp and Cook's (2005) main area of research was to develop a wireless, gesture-driven music controller which generates sound based on the physiological signals of the 'performer', thus removing the steep learning-curve required to master a musical instrument and instead creates a direct interface between emotion and sound-production. I believe this to be valuable, as from personal experience I realise that a musical performer can easily become overly absorbed in the mechanical aspects of instrumental control - often unrelated to emotional output - and sometimes forget about the emotion that should be conveyed by the performance. Any instrument that excludes this mechanical manipulation requirement would allow the musician the freedom to focus solely on the purpose of the music – the emotion he/she is trying to convey.

MES warrants investigation as the study at hand seeks to uncover methods for evoking emotion using noise-sound. I feel that a review of how MES achieves the 'musical effect' using musical sound may unveil some parallels to be explored using noise.

Juslin (2000) asks and attempts to answer two crucial questions with regards to the communication of emotion in musical performance: Can performers communicate emotions to listeners? What means do they use to accomplish this task? For answers he looks to 'performance expression', a field of study concerning "the small and large variations in timing, dynamics, timbre, and pitch that form the microstructure of a performance and differentiate it from another performance of the same music" (Palmer, 1997:118). His experiment involved using multiple guitar players who were asked to play the same melody numerous times – each time attempting to express a different emotion without altering the pitches of the melody. These various renditions were then auditioned by listeners in a controlled environment and listener feedback was given regarding whether or not the performers had indeed succeeded in conveying the emotions they had been asked to evoke. In most instances the listener responses indicated that the performers had in fact been highly effective in their given task. The resultant thinking is that performers must be using small deviations from the written melodies in order to evoke the different emotions. Carl Seashore said: "deviation from the exact . . , is the medium for the creation of the beautiful--for the conveying of emotion" (Seashore 1937:155). Seashore did not, however, suggest any theory to explain why these small deviations should evoke emotional reactions in listeners. It is perhaps for this reason that his statement did not give rise to further research on emotional expression in performance. Juslin (2000) highlights the following paradox in the literature in the field of music performance: Studies of emotional expression in music have almost entirely concerned themselves with the impact of particular pieces of music, whereas they have ignored the impact of specific performances. In addition performance studies have been largely concerned with structural realm of performance, and have ignored the emotional aspects. The result of this is that researchers know a lot about how different aspects of a musical composition might influence listeners' emotional responses to music but know less about how different performances might influence listeners' responses. It is clear that the same notated structure can be performed in many different ways, and the precise way it is performed has an influence on the listener's impression of the mood of the music. Therefore, it is important to study how the performance contributes to the emotional impact of music. In the results of Juslin's (2000) study, anger was associated – by both performers and listeners – with fast tempo, high sound level, a lot of high frequency energy in the spectrum, legato articulation, and small articulation variability; sadness was associated with slow tempo, low sound level, little high frequency energy in the spectrum, legato articulation, and small articulation variability; happiness was associated with fast tempo, high sound level, intermediate amount of high frequency energy in the spectrum, staccato articulation, and much articulation variability; fear was associated with slow tempo, very low sound level, little high frequency energy in the spectrum, staccato articulation, and large articulation variability. The correlations in the results show that the performers' expressive intentions had a considerable effect on the cues in the performances and that the cues in the performances had considerable effects on the listeners' judgements of the emotional expression. The differences in achievement among the three melodies performed were reasonably small. This suggests that the reliability of the emotional communication did not depend on the particular melody performed. In terms of validity for performers sound level had the highest value, followed by articulation, spectrum, articulation variability, and tempo. Interestingly for listeners the validities ranked differently: sound level had the highest value, followed by tempo, articulation, articulation variability, and spectrum. This indicates that there were some differences in the cue utilization among performers and listeners. Most notable was that listeners attributed greater importance to tempo than performers, whereas performers attributed greater importance to articulation than listeners. Juslin's study highlights 5 aspects of sound signals that his data suggests as being crucial for emotional communication in music: tempo, sound level, spectrum, articulation, and articulation variability.

Bresin and Friberg (1999) conducted similar music performance tests, however performances were synthesized from a musical score rather than auditioned through live recordings. In their tests - as in Juslin above – listeners were able to identify the intended emotions accurately. Musical cues they used to alter performances include some that are similar to Juslin's with a few added: tempo, sound level, articulation, attack and decay, timbre, and vibrato. As performances were synthesized in this study a set of 'rules' was written by which the software should perform each emotion.

One of the interesting future applications cited by these authors is the development of an 'emotion-toolbox' whereby users could select a piece of music and then have it interpreted by this software in the emotion of their choosing with the click of a few buttons.

Juslin and Laukka (2003) did a comparative study to determine whether the affective nature of music had anything to do with how we (humans) respond to speech cues – as we understand that the delivery of spoken-word definitely has an effect on the emotional impact it makes. They found many correlations between cues in speech and music. These are summarized below (from page 802):

Anger:

Fast speech rate/tempo

High voice intensity/sound level

Much voice intensity/sound level variability

Much high-frequency energy

High F0/pitch level

Much F0/pitch variability

rising F0/pitch contour

Fast voice onsets/tone attacks

Microstructural irregularity

Fear:

Fast speech rate/tempo

Low voice intensity/sound level (except in panic fear)

Much voice intensity/sound level variability

Little high-frequency energy

High F0/pitch level

Little F0/pitch variability

Rising F0/pitch contour

Lots of microstructural irregularity

Happiness:

Fast speech rate/tempo

Medium–high voice intensity/sound level

Medium high-frequency energy

High F0/pitch level

Much F0/pitch variability

Rising F0/pitch contour

Fast voice onsets/tone attacks

Very little microstructural regularity

Sadness:

Slow speech rate/tempo

Low voice intensity/sound level

Little voice intensity/sound level variability

Little high-frequency energy

Low F0/pitch level

Little F0/pitch variability

Falling F0/pitch contour

Slow voice onsets/tone attacks

Microstructural irregularity

Tenderness:

Slow speech rate/tempo

low voice intensity/sound level

little voice intensity/sound level variability

little high-frequency energy

low F0/pitch level

little F0/pitch variability

falling F0/pitch contours

slow voice onsets/tone attacks

microstructural regularity

Note: F0 = fundamental frequency

They also found that the importance of certain cues were more specific in music performance than in speech. According to their findings, the most important cue in music performance is articulation. Staccato articulation means that there is much air between the notes, whereas legato articulation means that the notes are played continuously. The results concerning articulation were relatively consistent. Anger, fear, and happiness were associated primarily with staccato articulation, whereas sadness and tenderness were associated primarily with legato articulation. One exception they found is that guitar players tended to play anger with legato articulation, suggesting that the code may not be entirely consistent across musical instruments. They noted that data regarding the use of vibrato were relatively inconsistent and suggested that music performers do not use vibrato to communicate particular emotions. Large vibrato utilised in anger portrayals and slow vibrato rate in sadness portrayals were the only consistent tendencies. Due to music's metrical nature, they also deemed it necessary to look at performances in terms of microstructural deviations from prescribed note values. Data concerning timing variability suggested that fear was the emotion where musicians use the most timing variability, followed by anger, sadness, and tenderness. Happiness demonstrations showed the least timing variability of all. Also, findings regarding durational contrasts between long and short notes indicate that the contrasts were increased (sharp) in anger and fear performances, whereas they were reduced (soft) in sadness and tenderness performances.

As part of their exhibition entitled 'ADA: Intelligent Space' (a physical environment with the ability to respond to the happenings in the space and create the ideal 'climate' of mood and emotion) at the Swiss National Exhibition Expo 2002, Wasserman et al. (2003) built a spontaneous music generator called the Roboser Music System. It is a digital system that accepts input from multiple sources to guide it in a real-time composition process. Roboser composes music on up to 12 performance tracks, the performance of every track being synthesized and performed in real time. Musical parameters that are interactively and independently controlled in each track include the MIDI parameters for instrument, velocity, volume, pitch bend, tempo, and articulation. In addition, predefined fragmented note sequences, rhythm lines, and note onset dynamic sequences are interactively selected for each performance track. Each track's output is delivered on a single MIDI channel. During the exhibition the outputs of the Roboser tracks were routed to a sampler, resulting in the creation of a complex soundscape. The creators of the ADA exhibit made a distinction between moods and emotions for processing of interactive features of the model, and as such Roboser's music performance was synthesized on two separate layers: mood and emotion.

Roboser's mood layer consisted of two mood parameters: arousal and valence (in accordance with Thayer's model as discussed in the MER section). The arousal parameter set the performances' tempo, volume, and octave register. As arousal at the inputs increased, the number of note onsets per second, note overlaps, and volumes increased while note pitches were shifted upward in octave steps. The valence parameter generated pitch material that changed from dissonance to consonance. Dissonance was expressed in semitone clusters for low valence. Consonant pitch material was taken from a harmonic series for high valence. For the emotion layer, they used a set of eight tracks to express the four emotions of joy, surprise, sadness, and anger. Two tracks were assigned to each emotion. With the onset of an emotion, the volumes of the respective two voices were gradually increased from zero to maximum, fading the emotion compositions in and out on top of the mood layer composition. The emotion compositions' musical features were based on an extension of the scheme outlined in Gabrielsson and Juslin (1996). However, the scheme was extended by the introduction of the emotion of surprise. Because of this extension the original scheme was altered to increase contrast between the four emotions. The sadness composition used slow tempo and mellow timbre sounds, and the scales comprised low registers and minor or diminished chords. To express joy major pentatonic scales, rather bright timbres, rhythmical lines, and moderately fast tempo were used. They represented anger with rough and distorted sounds, semitone scales, fast tempo, and mostly staccato articulation. Surprise entailed a very fast tempo, high volume, bright timbres, and mostly augmented chords.

3.4 CHAPTER SUMMARY

The above chapter began with a review of the goal of the current research. In order to answer this question, another question – what is music? – had to be dealt with first. Here various possibilities were explored, and ultimately a working definition for 'music' was synthesized. Music is 'organised sound which evokes an emotional response in the listener'. The question then became 'how does one experience the musical effect?' A review of the literature in this field seems to suggest that musical elements, working in conjunction with one-another, evoke what we call the 'musical effect' and move the listener on an emotional plane. These elements of music can be problematic as musicians have developed a system of labels for them which the musically untrained may not understand, yet they are still affected by music. An investigation was made into a reconciliation of the traditional musical elements and how the uninitiated hear music, and it was found that untrained people experience musical sound as a whole on an intellectual level, but later in the chapter we learn that our brains break music into its component parts even if we do not have a linguistic label for each part. For the purpose of this research the individual elements were identified by their traditional labels in order to explore possibilities in terms of their contribution to the musical effect – the listeners would not be exposed to these labels in the listening tests and, as such, the use of these labels should not pose a problem. The chapter then looks at how the brain goes about processing music, and the field of Strong Experiences in Music is explored. A list of common musical elements is uncovered and each element investigated for its affective strengths. The chapter concludes with an investigation into two more recent fields of music emotion research – Music Emotion Recognition and Music Emotion Synthesis, which both give some insight into how emotion can be identified in musical material.

CHAPTER 4 - LITERATURE REVIEW PART 2

4.1 DIGITAL SIGNAL PROCESSING TECHNIQUES

A signal is defined in the Collins English Dictionary (2003) as being '*any sign, gesture, token, etc., that serves to communicate information*'. An electronic signal is somewhat different in that it is called '*a variable parameter, such as a current or electromagnetic wave, by which information is conveyed through an electronic circuit, communications system, etc.*' A digital signal is different from both of the above in that it is a piece of information that has been encoded using the binary number system.

From the above definitions we can conclude that signals are representations of information and are vital to the successful transmission of information, given that one cannot transmit something that cannot be physically represented. Thus a signal is a sign that information is present in a system. However a signal may also be the means of transmission of the information it represents (electronic signal), or it may be an encoded representation of the information it represents (digital signal).

In this project we will attempt to manipulate a 'noise' signal – a digitally encoded representation of noise information – via digital techniques, to evoke a 'musical effect' (discussed in the previous chapter). Thus, one can see that signal should play a vital role in the process. Because this study will be conducted in the digital domain, digital signal will be the type of information input to be manipulated by the various techniques to be uncovered. Digital signal is electronic information composed of the binary digits 1 and 0 (called bits) used to represent 'on' or 'off' pulses. It differs from analogue signal in that digital signal is a stream of discrete values, whereas analogue signal is a continuously varying stream of infinitely possible values. Digital signals may be discrete-time representations of analogue signals (for example one may digitally record the sound of a musical instrument – the result being a digital signal representing that sound), or they may be created inside a digital system (for example sounds created in a digital synthesizer).

As stated above, the signal to be used in the listening experiments attached to this study will be a noise signal. Noise is a word that can be used to signify a number of different phenomena, and at this point it seems fitting that a firm understanding of the noise to be utilized here is discussed.

4.1.1 Noise and Noise Spectra

The word "noise" is derived from the Latin word "nausea," meaning seasickness (Weil 2008). Noise is commonly perceived as unwanted sound. Many authors describe noise as being any sound without musical quality or any unwanted or undesired sound (Hegarty 2001; Ruckman 2007; Sangild 2002; Ballou 2002).

From a scientific perspective, noise is simply sound with a continuous, non-periodic structure, or sound with all frequencies having random relative amplitude and phase (Chang 1999; Russ 1996; Roads et al. 1997; Martin 2011), which may possess many varied characteristics, such as "continuous or intermittent, random or semi-random, coloured (containing identifiable frequency components), impulsive, crackly, clicky, ticky (containing primarily high frequencies), or poppy (containing primarily low frequencies)" (Katz 2002). Noise has no harmonic structure, although it may be present only in specific parts of the spectrum (Russ 1996). In saying noise has no harmonic structure Russ (1996) refers to the idea that noise-sound may be considered the opposite of musical sound, where musical sound is inherently structured in terms of the frequencies (harmonics) it contains.

Noise is crucial to many inherently musical processes – the reason it cannot be excluded from musical signal processors that at first attempted to produce the sounds of musical instruments without the "noise elements" made in their playing. Attempts to exclude the 'noises' naturally produced in the production of music, via musical instruments, have resulted in sounds that are perceived as being unrealistic, or unnatural-sounding (Roads et al.1997; Russ 1996).

The specifications for white, pink, blue and black noise – the principal types in scientific use - are all found in The Federal Standard 1037C Telecommunications: Glossary of Telecommunication Terms.

4.1.1.1 White noise

White noise is defined as a noise that has an equal amount of energy or density per unit frequency (Hz). Because all frequencies have equal level, we call the noise white - just like light that contains all frequencies equally is white light.

This sounds 'bright' to us because we hear pitch in octaves. One octave is a doubling of frequency, therefore 100 Hz - 200 Hz is an octave, but 1000 Hz - 2000 Hz (not 1000 Hz - 1100 Hz) is also an octave. Since white noise contains equal energy per Hz, there is ten times as much energy in the 1 kHz octave than in the 100 Hz octave (Martin 2011:194).

White noise is the type of noise to be used in the listening tests of this study as it is the type which is spectrally the most even (see above). This evenness eliminates the possibility that test-subjects may be emotionally moved by the features inherent in any of the other types of noise and thus ensures to the greatest possible degree that subjects will be affected by the DSP techniques being used to manipulate the noise.

4.1.1.2 Pink noise

Pink noise, or $1/f$ noise, is noise that has a power spectral density which is inversely proportional to frequency, in other words an equal amount of noise power per octave. This means that there is 50% less energy per octave as you ascend in frequency – or, in more technical terms, power density decreases by 3dB per octave.

This is used because it sounds relatively 'equal' in distribution across frequency bands to us (Martin 2011:195).

4.1.1.3 Blue noise

Blue noise is the opposite of pink noise in that it increases by 50% (3dB) in power density each time you go up an octave (Martin 2011:196).

4.1.1.4 Red noise/Brown noise

Red noise is used when pink noise lacks the adequate low-end intensity. In the case of red noise, there is a 6dB drop in power for every doubling of frequency (Martin 2011:197).

4.1.1.5 Purple noise

Purple noise has an increase in power of 6dB for every increase in frequency of one octave (Martin

2011:198).

4.1.1.6 Black noise

Black noise is essentially silence with the occasional randomly-spaced spike (Martin 2011:199).

Due to the random periodicity of noise, the measurement of the spectral characteristics of noise is difficult as the noise's spectrum is constantly changing without following a consistent pattern of flux. In order to ensure accuracy when measuring, measurements must be taken over a long enough period that the noise is given sufficient chance to exhibit its full spectral character.

Moreover, the above definitions of the various types of noise are not entirely accurate in that they are based on the noise having a defined probability of having a specified spectral content. The degree of probability increases with the length of time for which measurement of the spectrum is taken (Martin 2011:200).

4.1.1.7 Iterated Ripple Noise (IRN)

Rippled Noise (RN) is created by delaying a random noise and adding it back to the original. Iterated Ripple Noise (IRN) is created by repeating this delay-and-add process. IRN produces a 2 component perception: a 'buzzy' tone with a pitch equal to the reciprocal of the delay and a background noise that sounds like the original random noise (Patterson et al. 1996:3286). The delay and add/subtract operations impart a spectral ripple and a temporal regularity to the noise (Yost et al. 1998:2349).With increasing delay-and-add cycles (or gain in each cycle) the stimulus becomes more regular, as shown by the increasing first peak in the autocorrelation spectrum (Griffiths 1999:136).

4.1.2 Digital Signal Processing

In order to better understand Digital Signal Processing (DSP) we must first understand its fundamental elements. Only then can a survey of DSP techniques be performed in order to uncover potential methods to be used in creating a 'musical effect'.

Before the advent of digital technology, the only way to store or transmit an audio signal was to use a means of recording a change in voltage as a waveform that was analogous to the pressure waveform that was the sound itself. This analogue signal recording method was successful, but suffered from the unwanted introduction of noise (Martin 2011:587). Due to the search for a better solution to representing recorded sounds more accurately we now have the system of converting analogue sounds to digital representations of those sounds – digital audio.

Digital audio is a means of using a series of discrete values to represent a sound pressure wave's change in amplitude over time. All digital operations are represented using the binary number system, allowing the use of fundamental arithmetic and logic operations (Pohlman 2000). Instead of recording the continuous change in voltage of a sound-wave (as in analogue recording), this digital process samples the voltage of the incoming wave multiple times per second and records the sampled values as binary numbers (at discrete time intervals), in order to represent the entire waveform as a string of numbers which could be plotted on a graph. Thus, in order to record sound (which is analogue in nature), digital systems use sampling and quantisation to transform the audio into a digital signal (Kefauver 2001:333; Roads 1996; Russ 2004:47). Sampling is the process of digitally capturing the sound-wave's amplitude – height in a transverse wave or compression and rarefraction in a longitudinal wave – a certain number of times per second. Sampling reduces a continuous curve to a series of discrete data points that exist at regular time intervals (Martin 2011:180; Ifeachor and Jervis 2002; Kefauver 2001:333; Russ 2004:47-50). Quantisation captures the wave's amplitude at the specified sampling rate in continuous time, and gives each measurement a digital value. When combined, a digital representation of the analogue sound is produced. Nyquist's sampling theorem states that for the result of sampling to remain true to the original signal, the sampling rate should be at least twice the highest frequency contained in the wave to be represented as digital data. (Kefauver 2001:333-335; Pohlman 2000:22,23; Russ 2004:47-50; Cook 2002:1-4).

The field known as Digital Signal Processing (DSP) spans the mathematics, algorithms, and techniques used to manipulate signals once they have been converted to digital form, thus DSP is concerned with representing a signal in digital form and using digital processing to analyze, modify, and extract information from the signal (Roads 1996; Ifeachor and Jervis 2002). DSP has been in existence since the 1960's. It was originally developed to process information for radar and sonar, oil exploration, space exploration and medical imaging. As personal computers became the norm, DSP began to expand its role into most areas of technology such as cellular telephones, CD players, and electronic voice mail. DSP has had such a diverse impact because it deals with such a standard format of information: digital signal. It cannot discriminate between audio or visual signal, for example, since to a computer, they both just look like a string of bits (ones and zeros).

Digital signal processing (DSP) has become a more accessible solution to employ in digital audio processing than is analogue signal processing, and has an emphasis on real-time operation (Roads 1996). DSP allows the user to “generate, analyse, or otherwise manipulate” digital signals. It employs the same means as digital recording (sampling and quantisation), but is used to process audio, rather than store it (Pohlman 2000:22,23).

Any DSP process can be defined by a mathematical algorithm that alters values in a digital bitstream (Martin 2011:663; Ifeachor and Jervis 2002; Pohlman 2000).

Some of the vast number of audio applications that DSP allows for are:

- Error correction
- Multiplexing
- Sample rate conversion
- Speech and music synthesis
- Data reduction and data compression
- Filtering
- Adaptive equalization
- Dynamic compression and expansion
- Reverberation, ambience processing
- Time alignment
- Acoustical noise cancellation
- Mixing and editing
- Encryption and watermarking
- Acoustical analysis

(Pohlman 2000; Roads 1996; Kefauver 2001; Russ 2004; Cook 2002; Case 2007).

One of the reasons digital audio is so important in today's music is that it maintains its integrity over a long physical distance (cable-run), and time (storage on CD or hard drive) – though some argue that we really have no idea how long digital storage media will hold information due to the field still being relatively young. Large numbers of digital copies can also be made without any loss of quality. DSP also allows music/sound engineers to work with many more tracks of music simultaneously than did analogue multitrack tape, and on one piece of hardware, as opposed to using multiple tape reels. It allows for programmed mix-down of vast numbers of tracks (signal addition and subtraction), signal filtering, and signal editing. DSP also enables the engineer to add one signal to another to create new sounds, and to manipulate a single signal via an algorithm to alter the original signal (e.g. addition of reverb). Some other basic functions of DSP include equalization (to modify the frequency spectrum of a signal), time-delay effects, convolution, noise reduction, and time compression/expansion (Roads 1996; Ifeachor and Jervis 2002). Although some of these features of DSP were available in the analogue domain, one of the major advantages of DSP over analogue is that all of these DSP functions can be performed on one machine, whereas, in analogue, multiple pieces of equipment would be needed to perform these duties. Also, many analogue techniques have become much more effective and diverse due to the increased processing power and flexibility offered by transferring these functions to DSP equipment.

Some of these primary uses of DSP (filtering, transformation, modulation) are originated largely through the use of digital filters. The other main branch of DSP – convolution/correlation – accounts for many other DSP techniques to be investigated (Ifeachor et al. 2002; Kefauver 2001; Pohlman 2000; Cook 2002).

4.1.3 Digital Filters

A digital filter is a computational process or algorithm by which a digital signal or sequence of numbers (acting as input) is transformed into a second sequence of numbers termed the output digital signal (Rabiner et al. 1972 in Roads 1996:397).

The definition above means that any digital device with an input and an output is a filter, and that any operation performed on a digital signal may be credited to a filter. However, the use of a filter usually involves some form of boost or attenuation (amplitude) or a shift (phase) of a portion of the spectrum of the signal in either the time or frequency domain (Ifeachor et al. 2002; Roads 1996:185).

4.1.3.1 Some advantages of digital filters over analogue filters

- Digital filters, being essentially software-based, are instantly reconfigurable and eminently more versatile than their hard-wired analogue counterparts.

- Digital filters can be designed by anyone with a personal computer and a basic understanding of music and the functions used in DSP, whereas analogue filter-design requires electronics knowledge and the skill to assemble the components of the filter.

- Digital filters are extremely stable with respect to time and temperature. This is one of the key features of DSP as it allows for the same result to be achieved by digital equipment irrespective of environmental conditions.

- Digital filter equipment is much more compact than analogue hardware and consumes less power than the analogue equivalents would.

- The performance of a digital filter is exactly repeatable from unit to unit, whereas analogue components rely on the integrity of the individual physical components that can and do sometimes affect performance.

(Ifeachor et al. 2002)

Ifeachor et al. (2002) discusses a few disadvantages of digital filters, such as limited processing speed, wordlength sometimes not being of high enough resolution to capture an accurate picture of the signal, and the time it takes engineers to design a filter. This text was printed some years ago and it is my feeling that these disadvantages have since been nullified through the advancement of technology.

4.1.4 Convolution

Convolution is one of the most basic operations in DSP. In fact, Ifeachor et al. (2002) defines digital filtering as the convolution of an input signal with a filter's impulse response in the time domain. The term 'convolution' refers to the 'mixing together' (by multiplication) of two unrelated values to create a product of them by multiplying every value in the signal with each value in the impulse. A filter convolves its impulse response with the input signal to create an output signal. Two properties that make convolution so useful are that it is shift-invariant and linear. Shift-invariance means that every part of a signal/information is acted on equally, and linearity speaks to the idea that each piece of the signal is replaced with a linear combination of its neighbours. These properties make convolution processes simple, as it is simple to do the same process at every point in a signal, and a linear task is also a simple one to process digitally (Jacobs 2005:1).

The Law of Convolution states that

Convolution in the time domain is equal to multiplication in the frequency domain and vice versa (Roads 1996:424; Cook 2002:56).

What follows is an overview of various filter types and the effects they apply to a signal.

4.1.5 Basic digital filters and their properties

The easiest way to characterize various types of filters is to examine their amplitude-versus-frequency response curve. As a graphic representation, a straight line parallel with the amplitude axis indicates a filter displaying no boost or attenuation across the spectrum – essentially a filter through which a signal may pass unaffected. Digital filters operate on streams of numbers sampled in time. Digital filters may make use of previous output values to provide new input information, sometimes by adding the previous output to the new input by feeding it back to the input stage. The linearity in a digital filter system (in which no new signals are created) allows for mathematical and processing techniques to analyze and predict system behaviour, and if a system is linear, Time Invariant impulse response can be measured and thus characterize its behaviour (Cook 2002:21-24; Roads 1996:186).

The cutoff frequency of a filter is the frequency at which the filter has reduced the signal to 0.707 of the input value. This is also known as the half-power point as 0.707 relates to 1 as equivalent to a 3 decibel (dB) decrease in audio terms. Anything below the cutoff frequency is said to be in the stop-band of the filter, whilst anything above the cutoff frequency is said to be in the pass-band in a high-pass filter – the opposite applies for a low-pass filter. These sounds in the stop band of a filter are attenuated according to a specific slope, usually measured in dB/octave. The specific musical situation will dictate how steep a slope would be used (Roads 1996:186-7).

In order to best observe a filter’s effects on a signal in the frequency domain, white noise is used as an input, as it contains equal proportions of all frequencies, whilst to observe a filter’s time domain response, we must use the shortest possible signal, one-sample in length. The output a filter generates from a one-sample input is called the Impulse Response of the filter. A filter’s impulse response directly corresponds to the amplitude vs. frequency curve of the filter, and is simply a time-domain representation of the frequency response of a filter. Phase response is a measure of how much a filter effects phase shift (in radians) of the component frequencies of a signal between its input and output. This can also be represented as phase delay (measured in seconds) in the time-domain (Pohlman 2000:94; Roads 1996:400-402).

The ‘order’ of a digital filter refers to a number which denotes the highest exponent in the numerator (feed-forward) or denominator (feed-back) in the z-domain of the transfer function of a digital filter. For FIR filters (discussed below), there is no denominator in the transfer function and the filter order is merely the number of coefficient/delay pairs used in the filter structure. For IIR filters (discussed below), the filter order is equal to the number of delay elements in the filter structure. It is common for complex filters to be constructed from many simple, efficient filters (first- and second-order) rather than building one complex one (Cook 2002:24; Roads 1996:412).

Digital filters can be put into two main categories: Finite Impulse Response (FIR also known as non-recursive), and Infinite Impulse Response (IIR also known as recursive). FIR filters calculate the current output from the current and previous input values and produce a finite number of outputs, whilst IIR filters use previous output values and current input values to calculate the current output thus yielding infinite outputs. IIR filters are also known as ‘feedback’, 'exponential decay', or ‘recursive’ filters. FIR filters are believed to be better for audio applications because they are easily designed with a linear phase response, thus preventing phase distortions in the audio output. FIR filters are also more stable than IIR filters because they have no ‘feedback’ component, whereas IIR filters tend to be excited by transient sounds causing a ‘ringing’ effect (oscillation of the filter). IIR filters are also more likely to compound errors in input signal as that input becomes an output with an error, and then that output is fed back into the filter as a new input with an error and so on. The only disadvantage of FIR filters is that they are much more processor intensive than IIR filters as the IIR filter’s automatic feedback of the earlier output values makes for less calculations to be performed by the processor (Cook 2002:24-26; Ifeachor et al. 2002; Martin 2011:687,694; Roads 1996:406-411).

The four most basic types of filter are: low-pass (allows only frequencies below a chosen threshold to pass through); high-pass (allows only frequencies above a chosen threshold to pass through); band-pass (allows sounds between two chosen frequencies to pass through); and notch (rejects sounds between two chosen frequencies). A shelving filter boosts or cuts all frequencies above or below a given threshold in equal amounts. Low-pass and high-pass filters are usually created using a first-order filter design as only one output is required (Cook 2002:30; Roads 1996:186).

A simple low-pass filter is called an averaging filter in that it averages the values of the current and previous input signals by adding them together then dividing by 2. It attenuates high frequencies by smoothing out the 'spikes' in the input signal (spikes are sudden changes or signals of higher frequency), and thus allows only lower frequency content to pass by unaffected. In order to create a more dramatic attenuation effect more previous inputs can be added to the current input and averaged (Case 2007:114; Roads 1996:403-404; Martin 2011:377).

Simple high-pass filters subtract samples to calculate the difference between them, rather than adding them together. It subtracts the previous sample value from the current one, and then divides by 2. This has exactly the opposite effect of a low-pass filter in that it exaggerates the 'spikes' previously discussed, rather than suppressing them (Case 2007:114; Roads 1996:403-404; Martin 2011:378).

In a band-pass filter the difference between the two threshold frequencies is called its bandwidth. Center frequency is the maximum point of boost in active band-pass filters, and the maximum point of attenuation in notch filters. This type of filter also makes use of a gain control, allowing for the amount of boost or attenuation to be decided by the user (Martin 2011:379-380; Roads 1996:187-8).

A comb filter is a filter comprising several regularly spaced sharp filter curves spaced evenly throughout the spectrum of the input signal. It derives its name from the visual qualities of its frequency response curve. FIR comb filters process their previous inputs. A FIR comb filter splits the input signal into two signal paths, and adds a tiny delay to one, then sums them again. This delay creates frequency-dependent phase cancellation and phase reinforcement when the two signals are summed, creating the comb effect. The delay time is manipulated in order to affect the desired frequencies. IIR comb filters operate in a similar manner, except that they process previous outputs (as opposed to inputs in FIR filters) (Roads 1996:194,412-417; Case 2007:231).

An all-pass filter is a filter that, ideally, would allow all frequencies to pass through it unmodified. The purpose of this type of filter is to apply phase shift to a signal without altering its spectrum. Since all filters introduce some amount of phase shift to the signals they process, an all-pass filter may be used to correct the phase shift applied by another filter. A simple chorus effect may be created using an all-pass filter, taking advantage of its inherent phase shift and delay (Pohlman 2000; Roads 1996:194-5,417-418; Martin 2011:707).

4.1.6 Filter banks and equalizers

A filter bank is a group of filters to which a signal is sent in parallel. In most filter bank applications, each filter is a narrow band-pass filter centered on a specific frequency, so as to allow for individual amounts of equalization (or spectrum shaping) across the frequency spectrum. The sounds are generally combined and routed to one output once they have been processed by the filter bank (Roads 1996:193-4; Cook 2002:76).

Equalization was first developed when recording equipment was not advanced enough to send an input signal to its output without having 'coloured' the signal on the way through – so it was used to make the output signal 'equal' to the input (Martin 2011:375; Case 2007:103). Technology has since moved on to a point where there is dramatically less unwanted alteration imposed on input signals in modern systems, however equalization is still used to make up for irregularities in frequency response of audio devices, or for more musical applications where a frequency, or frequency-band, may stand-out too much or be inaudible in a mix. In other words, we use equalization to modify the frequency spectrum of a signal (Case 2007:104). Common parameters on equalizers include: frequency select (the frequency to be altered or the frequency around which the signal will be altered); cut/boost (amount of gain or attenuation to be made); and bandwidth to be cut or boosted (increases or decreases the range of frequencies affected by cut/boost) (Case 2007:106).

All equalizers can be broken down into various types: a 'graphic' or 'multiband' equalizer splits audio into various fixed frequency bands and then boosts or attenuates those bands individually. A 'parametric' equalizer has fewer filters, and gives the option to select from a range of center frequencies around which equalization may take place. A 'notch' filter is used to cut a selected range of frequencies (usually with variable bandwidth), and a 'band-pass' filter eliminates all frequencies except the band it is set to allow through (Kefauver 2001:207; Roads 1996:194; Case 2007:106-109).

Some other interesting sound effects can be achieved with the use of equalizers. A simple high-frequency roll-off on a sound will cause the perception that the sound is more distant, as high frequencies diminish more over distance than low frequencies. Making a particular sound 'brighter' in one ear than the other (by making one side of the stereo image richer in high frequencies) will cause the listener to perceive that the sound is propagating from that direction as high frequencies are also attenuated by the human head whilst low frequencies refract around it better. Lastly a 'Wah-Wah' effect (common to guitarists) is created by simply boosting a frequency with a narrow bandwidth and then sweeping that boosted portion of the spectrum up and down the frequency range or using a LPF with sharp resonance and sweeping its center frequency (Case 2007:122-128).

Of course, boosting frequencies pushes them closer to maximum level, and if overdone may lead to distortion. Instead of boosting a frequency, another frequency can often be cut to achieve the same effect with less danger of causing distortion (Kefauver 2001:218; Case 2007:98; Pohlman 2000:70-90). Distortions in a digital signal are generally perceived as unwanted 'noises' – apart from in certain genres of music (such as heavy metal) where distortion plays a major role in creating the signature character associated with the music, where it can even be said to excite the listener (Case 2007:97). When distortion/clipping occurs, harmonics are added to the signal, altering the signals frequency spectrum without intention (Case 2007:92,98).

4.2 COMMON OPERATIONS USING DSP

4.2.1 Mixing

Mixing refers to the addition (positive or negative) of signals, and to the controlling of the relative levels of signals. This can be in terms of volume level, the level of certain frequencies in the signal (EQ), or the level of a signal on one side of a spatial field in relation to the other (Pan). Mixing can be performed using hardware or software applications, and performed as a real-time or non real-time process (Roads 1996:349).

Mixing can be viewed on two levels of meaning: The first is the purely physical or numerical meaning in terms of the relative levels (as values) of each of the component signals in the mix. The second level of meaning is the perceived mix (from a psychoacoustic viewpoint), which is also affected by spatial positioning of the signals, and effects such as masking which alter the final product as the audience hears it (Martin 2011:318).

4.2.2 Dynamic Range Processing

The dynamic range of a system or signal is the difference between the loudest and softest producible sound. Digital systems have a definite dynamic limit above which distortion occurs. The dynamic level in a system at any given time is calculated by the sum of all the sample values in the system at that moment (Martin 2011:408; Roads 1996:349).

Dynamic Range Processing is concerned with the alteration of the amplitude of signals through the use of various techniques such as:

- Envelope shapers
- Noise gates
- Compressors
- Limiters
- Expanders
- Noise reduction
- Companders

These techniques are usually used for their most practical function, but some of these also avail to the user some creative applications (Kefauver 2001:218; Case 2007:132; Roads 1996:390).

An envelope shaper is any tool used to rescale the shape of the overall amplitude of a given sound. This could be through the use of a gain change, or a complete redesign of the shape of the wave itself with regards to the amplitude vs. time curve.

A compressor is an amplifier whose gain is controlled by the signal level at its input (on some an external source can be used to control gain). A signal is accepted at the input, and if its level is detected (by a peak detector) to be above a desired ‘threshold’, it is attenuated to provide a more constant and even output level. The ‘transfer function’ of a compressor is a description of how a certain compressor setting would affect a given input signal.

Common compression parameters include: ‘ratio’ of change from input to output signal level; 'threshold' at which the compression will begin; 'attack' is time taken for the compressor to react to signals above the threshold and make 63% of the desired gain change; 'release' is length of time the compressor will act on a signal above the threshold before releasing it; 'make-up gain' is the amount of gain to be added to the compressed signal at the output (Martin 2011:408; Case 2007:132-135; Roads 1996:391-394).

The various implementations of compressors include digital compressors, tube compressors, optical compressors, voltage controlled amplifier compressors (Case 2007:137-138).

Compressors are usually used to reduce the dynamic range of a performance, however they may also be used in some more creative applications including to overcome noise; to increase perceived loudness; to improve articulation of vocals by using variable gain to increase volume of important syllables; to smooth an inconsistent performance; and to alter the envelope of a sound, thus changing the spectral shape of the sound (Case 2007:140-145). High compression ratios have been noted to create various effects such as the sustained ringing of the strings of plucked instruments, and the exaggeration of certain oral sounds in vocal performances. High ratios have also been noted to affect the timbre of certain instruments due to the reduction of natural transient sounds that are integral to the perception of specific timbre (Kefauver 2001:222; Roads 1996:391-394). Attack and release times also affect the perception of tone, as longer attack times, for instance, allow more high frequency transients to pass through than do short attack times. This feature of compressors allows the user to manipulate the compressor as if it were an equalizer to provide varying tones, rather than just a change in level (Roads 1996:391-394).

Another more extreme form of compression is found in a limiter. This is a compressor that, instead of rescaling an input signal above a threshold level, simply cuts it off at the threshold. This means that any input signal above the threshold is sent out at the threshold level. In DSP however, this cutting-off of the input signal usually means the production of a squared-off wave i.e. distortion in the output signal (Roads 1996:394; Case 2007:132).

An expander is the exact opposite of a compressor in its function. It takes an input signal below a given threshold and decreases its level by a decided amount. As it is, in essence, the inverse of a compressor, it is mainly used to increase the dynamic range of a signal. An upward-expander increases any signal level above the set threshold, whilst a downward-expander decreases signal level below the set threshold – essentially pushing the noise-floor down (Roads 1996:394; Case 2007:167-168).

Noise gates are used to clear up signals containing constant (unwanted) noise, for example electrical hiss/hum from noisy components in a system. Generally the noise level falls below that of the desired sound, so the noise gate is set only to open when the sound level reaches a point where the desired sound is present, and to stay closed when only noise is being transmitted. Thus, noise gates do not remove noise from the desired sound, but rather mute the noise at times when no desired sound is present (Roads 1996:390).

The operations discussed above provide the basic framework for digital audio engineering to be possible. However, as in all art-forms, there are additional facets to DSP which make the field that much more diverse, and enhance the ability of the audio engineer to be creative in his work. Digital sound effects make use of the principles of DSP in order to do exactly that.

4.2.3 Delay

Before the advent of DSP, an audio engineer would have to follow a rigorous process in order to delay a signal and then use it again in some way. This involved the use of multiple tape machines playing back copies of the signal at precise times. In the digital domain, however, the process of creating such effects is much simpler. With the advent of DSP a digital delay line is used to manipulate the time-axis of a signal by making copies of the initial signal, and then holding the copies and releasing them at the desired time-intervals (Pohlman 2000:631,632; Case 2007:207). Case (2007) mentions various types of digital delay units, but most are built on the 'sample and hold' principle so it seems unnecessary to delve into these various types.

Delay units generally feature a number of parameters to be set by the user. The most basic of these are input and output level controls; delay time control; and a regeneration control which sends some of the output back to the input in order to add further delays to the signal, which is how an echo is made to repeat more than once. Delay time is the time interval between original sound and first echo of the original sound (Kefauver 2001:188; Case 2007:210). Delay effects are often made more complex through the addition of 'modulation' parameters. These are used to vary the delay and create fluctuations in the effect rather than the unit maintaining a constant delay time. These parameters include rate (how fast the delay time is changed); depth (how much the delay time is changed); and shape (the pattern of sweep of the delay time whilst it changes according to the rate function) (Case 2007:211,212).

Delays are classified according to their lengths (not all sources agree on exact values): long (greater than about 60 milliseconds); medium (about 20 – 60 milliseconds); and short (less than 20 milliseconds) (Dennis 2001; Case 2007:214).

4.2.3.1 Long delays:

Some examples of long delay effects are echo (multiple repetitions of a sound fading in intensity over time) and slap-back (a single, quick echo). Adding a long delay often adds emphasis to a sound or word, making it seem more important to the listener (Case 2007:225; Kefauver 2001:188). A musically timed long delay may also be used to add groove/rhythm to a song, as though it were part of the rhythm section (Case 2007:226).

4.2.3.2 Short delays:

A short delay is not perceived as an echo. We perceive one single sound which is a fusion of the original or direct sound and the short delay, this fused sound having a different sonic quality to the pure direct sound. Very short delays have a spectral effect on the direct sound through the phenomena of constructive and destructive interference – components of the direct sound which are in phase with the short delay are strengthened whilst parts out of phase are degraded causing the resultant sound to possess different frequency content than the direct sound. Together these two effects create a comb filter, an effect sometimes used in music production to emphasize particular frequencies (Dennis 2001; Case 2007:227,228,232).

Short delay differences across the stereo spectrum (left and right channels) also affect the perception of the direction from which the sound originates (via the Haas effect), and thus may be used to alter the stereo balance in a mix (Dennis 2001).

The 'flanging' effect was initially created by feeding the same sound into two separate tape machines simultaneously, and then manipulating the speed of one machine by applying pressure to the 'flange' of the tape reel, and combining the outputs of the two machine. The effect of this is cancellation and reinforcement of certain frequencies as the tape speed varies. Digital flanging is created using a digital delay line which is constantly shifting in speed between around 200 microseconds – 15 milliseconds (basically a time-varying comb-filter). The effect can be described as 'ringing, whooshing, ear tickling' (Kefauver 2001:202; Case 2007:239).

4.2.3.3 Medium delays:

A 'chorus' effect is created through the use of medium delay times on multiple copies of a signal which are very slightly detuned from one another to create the effect of many sound sources in place of just one (Kefauver 2001:204; Case 2007:243,244; Dennis 2001).

4.2.3.4 Varying delay:

The Doppler Effect can also be replicated through the use of the delay effect. This is the change in frequency that occurs to a delayed signal when the delay time changes as a waveform is formed. For example, as a car drives away from you, the pitch of the engine gets lower because the wavelength is lengthened by the car’s away-movement. Thus the end of the cycle is delayed longer than the beginning of the waveform giving the effect of time being added to the cycle. If the car were moving toward one, the wavelength would be shortened due to the car compressing the wavelength in moving closer to the listener. This can be recreated through the implementation of a steadily increasing or decreasing delay time applied to a given pitch (Dennis 2001).

Reverberation (reverb) is the effect of audible repetitions of a sound diminishing over time whilst occurring more and more frequently. The principle use of this effect in nature is that it provides information in our perception of our surroundings (Kefauver 2001:188; Case 2007:264). The decay or reverb-time (RT60) of a space is the time in seconds taken for the reverberant tail of the sound to die away by 60 decibels (Kefauver 2001:188; Case 2007:265). Predelay is the amount of time between the audition of the original sound and its first reflections (Case 2007:267). The popular 'doubling' effect used in music production can be achieved through the addition of reverb with a reverb time so short that the first reverberation is perceived to be inseparable from the original sound (Kefauver 2001:200).

Reverb is usually used to add a sense of space, sonic colour or texture to a sound source (Case 2007:307-312).

4.3 SYNTHESIS

Chambers 21st Century Dictionary says that synthesis is 'building up; putting together; making a whole out of parts' (Russ 2004:3). We understand from this that synthesized sounds must be sounds made up of multiple other sounds. Synthesis has been used for a number of years in music and has made great strides in terms of how well it can mimic real (or natural) sounds. Many types of sound-producing systems can be modeled in real-time due to technological advances (Cook 2002:xi). It should be noted that synthesis is not exclusively a means to imitate real-world instruments. Rather, one of its greatest strengths is that it provides us with the facility to create something new and unique.

Russ (2004:4) says that 'used carefully it can produce emotional performances which paint sonic landscapes with a rich and huge set of timbres, limited only by the imagination and the knowledge of the creator.'

Fourier analysis and Fourier transform are methods for solving for the various sine and cosine components of a complex wave (Cook 2002:53). If both time and frequency are variable and both must be sampled then Discrete Fourier Transform must be implemented – whereby each domain is processed individually (Cook 2002:54).

Fast Fourier Transform is an economical method of computing the components of a complex sound (sound containing many sinusoidal modes), but in order to make use of this technique the signal length must be a power of 2 (Cook 2002:57). FFT can detect approximate pitch by measuring frequency through the decomposition of the signal into its sine components and finding the fundamental (or the greatest common divisor of the signals peaks) (Cook 2002:58). The first 5-7 harmonics and fundamental are most important in pitch recognition for both human listeners and computational purposes (Cook 2002:58).

Short Time Fourier Transform is another transform method which breaks a signal into segments which can then be manipulated. It breaks any signal into small parts, the size of which can be set via 'window size' (length of segment) and 'hop size' (where each window is placed in signal) and can be manipulated to give an approximation of human audio perception (Cook 2002:60).

4.3.1 Synthesis using Filters

Filters may be built as fixed (parameters remain the same) or time-variant (parameters may be altered with the passage of time) (Roads 1996:196). Applications such as equalization mostly call for a fixed filter, as once the desired effect is realized the filter is set and left to do its job. In the field of subtractive synthesis (to be discussed hereafter), however, time-variant filters have been exploited to achieve some remarkable results (largely in the fields of dance music and speech synthesis).

4.3.1.1 Subtractive Synthesis

Subtractive synthesis occurs when a complex sound (e.g. noise) is used to excite resonant filters (Cook 2002:85). For example, a periodic constant impulse run through a filter with a particular set of parameters results in impulses being outputted with same spectrum as the parameters of the filter (Cook 2002:86).

The channel vocoder illustrates subtractive synthesis techniques perfectly. The first stage of the vocoder is a filter bank (fixed frequency bandpass filters) spread over the bandwidth. Each filter’s output is connected to an envelope detector generating a voltage based on the amount of signal in the filter. The second stage of the vocoder is another bank of filters, to which an identical input signal may be sent, the output of each being sent to a voltage controlled amplifier. These outputs are then all combined to a single output signal. The first stage produces a control signal that determines the amplitude of the signal coming from the second stage.

In musical applications, and because the control function is separate from the second stage of filters, one element of the excitation signal (such as timbre, pitch or rhythm) is individually alterable independently of the elements of the driving function, or vice versa (Cook 2002:75-80; Roads 1996:198-199).

Cross-synthesis generally refers to the use of the control function (pitch and event timing) of one signal to modulate another signal. This usually means that a certain amount of spectrum transformation will take place. A cross-synthesizer can be used to make non speech sounds 'sing' or 'talk' (Cook 2002:80). By feeding a human voice into the analysis filterbank and another sound-producer (e.g. a guitar) into the synthesis filterbank, the effect of a singing or talking guitar is created (Cook 2002:80).

A phase vocoder can shift time and pitch, cross-synthesize, and create the sounds of hybrid instruments. A phase vocoder can be used for cross synthesis in the same way as a channel vocoder. The analysis signal is analyzed for time-varying magnitude and multiplied by its spectral frames. It can also be used for time and pitch-shifting (Cook 2002:82). In a phase vocoder Fast Fourier Transform is used to process the sound by calculating and maintaining magnitude and phase, and because it does not assume that the input signal is vocal (speech) it may be used to cross-synthesize many other types of sound source, like a plucked trumpet for example (Cook 2002:81).

In Linear Predictive Coding (a method of subtractive analysis/resynthesis in which future values of a signal are predicted based on previous values) the (usually) simple control function of a signal is replaced with a more complex time-varying signal, and used to modulate the second signal (e.g. using the sound of an orchestra as control function to modulate a speaking voice). LPC synthesis can use this same method to “clone” like-sounding instruments of different pitch ranges, using the excitation function of one as a control frequency, and via analysis of its characteristics, one can then create the same effect in different ranges (Roads 1996:208-209; Cook 2002:88). The human auditory system is more sensitive to peaks than cuts in a signal. Since Linear Predictive Coding is more accurate in predicting future peak-values than averages, it is a convincing method for sound creation (Cook 2002:91).

4.3.1.2 Spectral Modeling and Additive Synthesis

In additive synthesis signals are added together to make new ones. In order to add signals and then make them sound realistic the spectrum of the signal one hopes to imitate must be understood and accurately modeled (Russ 2004:9).

Some portion of the harmonic spectrum is exhibited by all signals. Signals which are more periodic and have frequency content which is spaced at multiples of the fundamental exhibit a strong sense of pitch. We call these 'musical tones' (Cook 2002:63). So we use the fundamental and 'harmonics' to determine the pitch of the sound, but the same pitch may sound 'different' due to different forments, like the same note played by violin and trumpet does not sound 'the same' (Cook 2002:64). Humans use the spectral shapes of the sounds we hear to allow us to differentiate the producers of two sounds of the same pitch and loudness. Some of the characteristics we perceive are attack, sustain, harmonicity vs. inharmonic, and amount of noise inherent in the sound (Cook 2002:65).

Noise can be thought of as being any sound with no clearly apparent sinusoidal modes (Cook 2002:66), although those sinusoidal modes may be present in the noise.

Fourier Transform can show us which parts of a sound can be modeled using sinusoids and which parts must be created through the spectral shaping of noise for successful additive synthesis to take place (Cook 2002:68). The sinusoidal components of the sound are said to be deterministic, the noise components are 'stochastic' or random (Cook 2002:69). These filtered noise components are often added to synthesized sounds to add 'realism' as most sounds in nature have some component of noise inherent in them (Cook 2002:69). Transient sounds (or sounds too short to be analysed) are added by Pulse Code Modulation sample (Cook 2002:71). Pulse Code Modulation uses multiple recordings, filters and various interpolation methods to make synthesized sounds more 'real' (Cook 2002:xi).

4.4 CHAPTER SUMMARY

In this chapter an operating definition was given for 'signal'. As this study deals with noise signals the phenomenon of noise was identified and its features discussed. White noise was given as the chosen noise type to be implemented in the experimental portion of this research and the reason for this was supplied. The field of Digital Signal Processing was then investigated. Various DSP operations and techniques were discussed, ending in a review of basic synthesis techniques.

CHAPTER 5 - EXPERIMENT DESIGN

5.1 INTRODUCTION

This project hopes to determine whether or not a relationship exists between certain musical elements and certain digital signal processing techniques so that the musical effect may be evoked by means of the implementation of the selected techniques to manipulate a noise-signal. In order to be considered successful, the presence of the musical effect should be evoked in neutral listeners of varying levels of musical experience. In order to determine the presence of the effect and its evocative power, a sample group of unspecified musical training was selected and subjected to a listening test, whilst answering a questionnaire regarding the material they were auditioning. Questions were directed towards the respondents' emotional state at the points in the test at which the noise-signal is being manipulated. This was to determine whether or not these manipulations cause a change in the emotion of the listener because this, more than the various labels one could attach to the musical elements, is ‘the musical effect' (as discussed in Chapter 3).

Some of the issues regarding experimental design such as ethical considerations, sampling method and sample size, and the reliability and validity of the research have been discussed in Chapter 2, and thus, to avoid redundancy, will not be discussed again.

To follow is a summarized account of the procedure followed from the beginning of the implementation of the tests to their completion. After this summary certain issues requiring further attention will be discussed in greater detail.

The first step in beginning the testing-phase of this study was to assemble a sample group on whom the listening tests could be conducted. The methods used here have been discussed in Chapter 2 of this treatise. Once the sample group had been gathered, the general process for the administration of the listening tests was as follows. The test venue, the office of Andrew Warneke at Harvest Christian School in Walmer, Port Elizabeth, was prepared, and testing equipment was deposited in the venue. Care was taken to ensure that there were no distracting environmental factors such as extreme temperatures or excessive background noises, to ensure the comfort and thus the focus of the respondents. The venue was closed-off to avoid anyone accessing the venue during testing and thereby causing a disturbance. Each respondent entered the test venue and took his/her seat. Respondents were seated in a comfortable office-type chair at a desk in front of the CD player. The respondent was given an introductory letter and an informed consent form (both to be found in the appendix to this treatise) to read and complete. Each respondent was also welcomed and introduced to the project by me and an explanation and description of the nature of the tests similar to that in the introductory letter was given verbally in order to ensure that each respondent was aware of what he/she was to take part in. The respondent was then assured that he/she had the right to leave at any time if he/she felt any discomfort or no-longer wished to take part in the listening test. The respondent was made aware that the listening tests were perceptual in nature and that, as such, there were no right or wrong answers. The respondent was also assured of anonymity as names were not required on the questionnaires so neither I nor anyone else would know which response form belonged to which respondent. This was done to help respondents to feel more at ease when answering and to remove any fear of failure or public humiliation.

Once these preliminary steps had been taken the listening tests were conducted. First, a questionnaire was distributed to the subject. He/she was asked to read through the questions. The rating-scale used on the questionnaire was explained to him/her, and he/she was told that the test included some unaltered – or unprocessed – noise signals as control elements to measure response against. He/she was notified that they would be alerted by audio on the testing tracks as to when he/she should answer each question. The listening test was then conducted while the respondent filled in his/her questionnaire. At the conclusion of the test the questionnaire was collected and the subject was debriefed and asked if there were any questions regarding the listening test or the research being conducted. The subject was then thanked for his/her participation and allowed to leave the testing venue.

5.1.1 Venue and Audio Equipment used to conduct listening test

One of the major considerations when conducting a listening test is that subjects can only make accurate judgements of audible material if the material is reproduced accurately and clearly by the audio system they are listening on. With this in mind, I felt that it would be beneficial to use a high quality CD player, headphone amplifier and a pair of studio-reference headphones to transmit audio to the respondent. The use of headphones was chosen because the alternative option, stereo loudspeakers, does not assist in the task of isolating the respondent from background noise quite as well as headphones do.

CD Player:	TEAC PD-155mkII
Headphone Amplifier:	Behringer Powerplay Pro XL (HA-4700)
Headphones:	Audio Technica ATH M-20

After a thorough process of investigation into the functionalities of various software programs at my disposal and with the knowledge of which DSP operations needed to be performed for the creation of the listening tests, I decided to utilize the WAVELAB and REASON software packages as they offer a fairly comprehensive, powerful solution to DSP functions and were readily available and well-known to me. Each part of the test was pre-recorded in real-time at the NMMU Music Technology Studio 2, as the live performance of these techniques in the testing venue – although possible – may have lead to inexact replications of the test and, as such, skewed results. Pre-recorded test elements allow for more easily repeatable tests, and are more reliable as multiple copies of the tests can be produced on CD so that backups are available in the event of a faulty disc. Also, as the test was recorded in real-time, the real-time focus of the project is not placed in jeopardy.

In order to allow for this testing to be replicated if necessary, I chose to set the test material to play at 75dB SPL at the beginning of the test track. The use of headphones presented a problem in terms of measuring this nominal level as an SPL meter does not have the functionality needed to measure the loudness of a pair of headphones. However a suitable solution was found by the following process. I played the listening test material on my high end hifi equipment at home and measured a 75dB SPL sound level. I then auditioned the test material via the exact equipment to be used for the listening test and, to the best of my ability, recreated the feeling of the 75dB SPL measured on the hifi. The volume level on the headphone amplifier was noted and this volume was used throughout testing. 75dB SPL was selected as a nominal volume as it is a safe listening level and allows for wide volume modulations before reaching an unsafe loudness.

5.1.2 Musical elements to be tested via DSP operations

During the literature study portion of this treatise – Chapters 3 and 4 – I began to uncover and intuit various possibilities in terms of relationships that may exist between certain of the musical elements and DSP operations discussed. With these in mind a period of private experimentation was undertaken in order to test these possible links and unearth the most promising ones in terms of the goals of the project at hand. The techniques and effects to be tested in the listening tests are the ones that emerged as having the most potential for use in real-world applications, as they are the ones which seem to be both viable in terms of implementation in the digital domain and could be useful to musicians and composers.

It should be noted that in order to avoid the respondents becoming too acclimatised to each technique and thereby distorting results via a learning effect, wherever a portion of the test contains multiple components, these were split-up across the experiment rather than performed back-to-back.

Dynamics – as per Chapter 3 of this treatise – is the flow of change of volume and intensity of the musical information being performed. These changes are usually indicated on the score by the composer. The dynamics portions of the test will measure the difference in emotional response generated by the use of large and small dynamic ranges.

Due to the idea that the decibel scale is logarithmic in nature and that perceptually the number of dB's of change is not simply proportional to the perceived loudness (difference of 10 dB is an actual change in intensity of a factor of 10, and a difference in perceived intensity of about a factor of 2), 'large' and 'small' dynamic range must be qualified. For this test it was decided that 'small' dynamic change would be any change within 6dB from nominal volume-level - 75dB (as this is considerably less than a perceived doubling), and a 'large' change will be greater than 10dB from nominal volume-level (this is perceived doubling and more). The test will consider both increase and decrease in volume from nominal level. The nature of the fluctuation of volume change was also tested. 'Abrupt' volume change was implemented through sudden volume change (reminiscent of a square-wave), whereas 'continuous' volume change employs a more gentle (sine-wave like) curve. These dynamic changes were implemented through the use of a white noise sample created in WAVELAB and played from the ReDrum unit in Reason. The volume was modulated by the automation of a volume control.

Four dynamic effects were tested:

- small range; discrete change
- small range; continuous change
- large range; discrete change
- large range; continuous change

Spectral Shape is a combination of timbre (tone-colour) and register (pitch-height). These have been combined as, although I have uncovered much information that points towards these being important individual elements of music, noise-manipulation techniques do not seem to facilitate the creation of these individual elements. However, as per Kim (2010), Cook (2002) and van de Laar (2006), the spectral shape of a sound does have much significance in terms of our perception of the nature of the sound.

Alteration of spectral shape was conducted through the use of a white noise sample created in WAVELAB and played from the ReDrum unit in Reason. The noise was manipulated by the control of a shelving filter.

Four variants of spectral shape were included in the test. 'Bright' refers to high frequency energy above 2kHZ being enhanced by 10dB. 'Dark' refers to low frequency energy below 200HZ being enhanced by 10dB. 'Mid-heavy' refers to mid-frequency energy from 200 Hz to 2kHz being enhanced by 10dB. 'Mid-lite' refers to mid-frequency energy from 200Hz to 2kHz being decreased by 10dB.

Spectral shape effects tested:

- bright
- dark
- mid-heavy
- mid-lite

Rhythm, according to Thul (2008), is the description of the duration and durational patternings of musical notes. In his definition of rhythm, the term duration is used for the number of time units between musical notes and the fundamental (and indivisible) time unit to be the pulse.

Rhythms are differentiated through their complexity values, a concept first researched by Lerdahl and Jackendoff in their authoritative text 'A Generative Theory of Tonal Music' (1983). Their work has since been expanded upon by other researchers, one of whom is Toussaint. His Metrical Complexity Measure proposed in his texts of 2002 and 2007 calculates a sum of the metric weights for each onset in a given rhythm and assigns a complexity value based on that sum – the lower the value the more complex the rhythm. He arrived at these values via a fairly complex mathematical process which lies beyond the scope of this discussion. Suffice it to say that the results of his method have been successful enough to warrant their use in this test (Thul 2008).

Toussaint's formula for rhythmic complexity assigns a numeric value to each pulse in a given bar. A 4/4 bar is said to have 4 beats and 16 pulses. Below is a graphic representation of the values he assigned to each pulse in a bar.

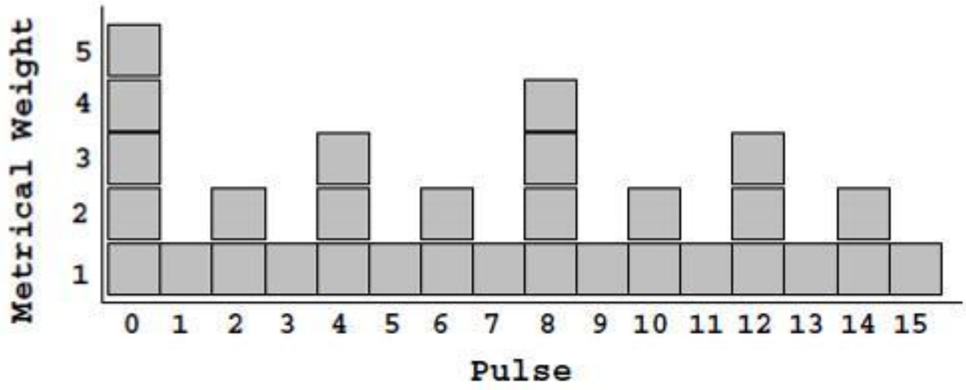


Figure 1: Toussaint's Metrical Complexity Measure for a bar of 4 beats

In order to test the evocative effects of rhythm, 3 rhythms of varying complexity were generated. A 'simple' rhythm with a complexity measure of 25, a 'medium' rhythm with a complexity measure of 17, and a 'complex' rhythm with a complexity measure of 10.

Below is a representation of these rhythms in traditional music notation.

Simple:

Medium:

Complex:

Figure 2: Examples of simple, medium and complex rhythms to be implemented in the listening experiment

Rhythms were generated through the use of a white noise sample created in WAVELAB and played from the ReDrum unit in Reason. Initially rhythms were to be generated via volume change, however volume change is not necessary for rhythmic perception as a pattern is perceptible whether volume change or not, and I felt that volume change might be perceived as dynamic change rather than rhythm. As such it was decided that the noise would be manipulated by the control of a band-pass filter, which creates the sensation of pattern without volume alteration.

3 Rhythm patterns were tested:

- Simple
- Medium
- Complex

Peak sound level is the highest level that is achieved during a piece of music. As per Chapter 3 of this treatise, peak sound level has been used as a means for determining the emotional content of musical material in the MER and MES fields, and is a suitable candidate for testing in this experiment as sound level is a parameter of a noise signal that can be manipulated via a fairly simple process.

The test of the evocative nature of peak sound level was administered by the increase in sound level of a portion of the noise signal by 30dB (from 65dB SPL to 95dB SPL) and then a decrease back to 65dB SPL. For this particular portion of the test the nominal level was reduced to 65dB SPL as a 30dB increase from 75dB SPL takes the level to 105dB SPL. Levels over 85dB SPL are legally unsafe over prolonged periods of exposure and in order to preserve the comfort and safety of respondents it was decided that 105dB SPL was too extreme as a peak-level, especially in headphones.

Four variants of this process were tested in order to gain understanding of whether or not time had a part to play in the effect created by peak sound level perception. A slow increase – decrease over 15 seconds, a medium increase – decrease over 10 seconds, a more rapid increase – decrease over 5 seconds, and a sudden increase – decrease over 1 second were all implemented in the test to provide a variety of values with regard to the time factor.

These effects were generated through the use of a white noise sample created in WAVELAB and played from the ReDrum unit in Reason. This noise was modulated in volume through the use of a volume control whose automation was timed to achieve the two desired cycles.

4 Peak sound level tests:

- 30dB slow (15 seconds)
- 30dB medium (10 seconds)
- 30dB fast (5 seconds)
- 30dB sudden (1 second)

Repetition is a very common musical device and Chase (2006) and van de Laar (2006) – both discussed in Chapter 3 – show that it is important in our perception of the musical effect. This part of the experiment was an interesting one, since because of the random and unstructured nature of noise a repetition of a part of a noise signal may or may not be perceptible to listeners. If it is, this may say something quite novel about our level of awareness when it comes to hearing random patterns.

For this part of the experiment a short (2 second) segment of our original white noise signal was cut out and then repeated 16 times to create a continuous 32 seconds of noise. This effect was created in WAVELAB as REASON was not needed.

6 Repetition Tests:

- 125ms
- 250ms
- 500ms
- 750ms
- 1 second
- 2 seconds

5.1.3 Questionnaire format and style

The goal of the listening tests was to determine whether or not subjects could identify a difference in evocative effect between various noise-signals that had been altered using various DSP techniques (described above).

The tests included only a perceptual element and not a subjective one, as subjects would not make decisions as to their preferences, but only with regard to the attributes contained in the signals they audition in the tests. One could also refer to the nature of these tests as 'non-hedonic' (Martin 2011), as subjects would not be questioned regarding their preferences for the materials they were auditioning.

Each question was presented using the Three Alternative Forced Choice (3AFC) method. Here the subjects are given three options to choose from as an answer to each test section. This was decided upon as it is one of the more reliable methods of testing when one is looking to exclude the subject's ability to 'score' well on a perceptual test by guessing. In being asked to choose one answer from three a subject has only a 33,3% chance of guessing correctly as opposed to a 50% chance in tests employing the ABX method (Martin 2011). A random distribution across a subject's answers would indicate that the subject had been guessing the answers or not taking the test seriously.

Thayer’s model of emotion (discussed in Chapter 3) was chosen as the most appropriate means with which to evaluate responses to the questionnaire in this study. No claim is being made here regarding its supremacy over other emotion-mapping models, however its use does offer some key advantages to the current research. The main benefit from the use of this model is the consistency generated as numerous other studies in related fields such as MER and MES (cited in Chapter 3) have been conducted using Thayer’s model. Another advantage it provides is in dealing with one of the problems encountered during the design of this listening test, which was the use of language-tags as descriptors for the various emotions which could be generated by the test materials, as it is impossible to say that every respondent would ascribe the same meaning to each tag used. Various studies have been conducted using variants of Thayer's model which have been altered to attempt to exclude language, or at least to simplify it. Tolos et al. (2005) found that the most universally applicable descriptors for testing in circumstances similar to the ones at hand are 'aggressive', 'happy', and 'calm/melancholic'. 'Aggressive' would represent high arousal – low valence; 'happy' represents high arousal – high valence; and calm/melancholic represents all low arousal states, as they found results in these quadrants of Thayer's grid to be the least accurate. They also conclude that the arousal axis of Thayer's model is easier to define than the valence (positive or negative) axis for most people. This was the model which was selected for the questionnaires.

At each point in the experiment where a respondent would need to provide information on the questionnaire, he/she would simply underline/circle the descriptor which he/she felt most described the effect generated by the material being presented.

So each info-point on the questionnaire would be depicted as follows:

happy aggressive calm/melancholic

As stated above, respondents would be alerted to the need to answer a section on the questionnaire by a voice-prompt in the audio attached to the test sounds.

The test elements were set-out in the following order so as to avoid a learning-effect compromising results:

1. Dynamics – small range; continuous change

2. Spectral Shape - dark

3. Rhythm - medium
4. Peak Sound Level - medium
5. Repetition - 2 seconds
6. Rhythm - simple
7. Spectral Shape - mid heavy
8. Peak Level - sudden
9. Dynamics – large range; discrete change
10. Repetition - 500ms
11. Repetition - 1 second
12. Peak Sound Level - slow
13. Dynamics – large range; continuous change
14. Rhythm - complex
15. Spectral Shape - bright
16. Repetition - 125ms
17. Peak Level - fast
18. Repetition - 750ms
19. Dynamics – small range; discrete change
20. Spectral Shape - mid lite
21. Repetition – 250ms

5.2 CHAPTER SUMMARY

In order to determine whether or not noise can be manipulated to evoke the musical effect, data must be generated. A listening test was chosen as the appropriate method to be used to gather this data. In this chapter the process of administration of the listening experiment was discussed. Items such as test venue, equipment used, and test material preparation were described and the procedure followed was detailed. The logic and format of the listening test questionnaire was presented, as were the musical elements to be utilized in the creation of the audio material. The chapter concludes at the point of the actual implementation of the listening tests, and Chapter 6 (to follow) will seek to collate and analyze the data. Thereafter, in the final chapter, a discussion of the results will be given and conclusions drawn.

CHAPTER 6 - DATA ANALYSIS

6.1 INTRODUCTION

The focus of this chapter is on the presentation and analysis of the results obtained in the listening experiments described in the previous chapter. The main and specific aims of the study – discussed in Chapter 1 but restated here, are to ascertain whether the musical effect can be evoked through the digital manipulation of a noise signal, and then to possibly identify a technique or various techniques that can be used to control noise in real-time to evoke the musical effect. These techniques should be practically reproducible with some training in their implementation.

The specific aims of the study can be summarized through the following four questions:

1. Is it possible to evoke the musical effect using real-time noise control methods?

2. Which musical elements are the most evocative of the musical effect when implemented via a noise-control technique?

3. Which properties of digital processing methods are best suited to be used in the creation of noise signals which evoke the musical effect?

4. What are the possible applications of these techniques in both the music world, and other spheres?

The chapter will present, as descriptive statistics, the consolidated results of the listening experiments so as to begin to provide insight to the possible answers to the questions above.

The data-set collected from the listening tests consists of the responses of 31 test-subjects to a 21 question questionnaire which was answered while listening to the test material in the testing venue described in Chapter 5. Subjects were asked to rate each piece of manipulated noise-signal in the test as either happy, aggressive, or calm/melancholic in keeping with the rating scale implemented in the work of Tolos et al. (2005) who found these to be the most universally acceptable language labels for a test based around the Thayer model of affect. 'Aggressive' would represent high arousal – low valence; 'happy' represents high arousal – high valence; and calm/melancholic represents all low arousal states as they found results in these quadrants of Thayer's grid to be the least accurate. They also conclude that the arousal axis of Thayer's model is easier to define than the valence axis for most people. Subjects also filled in informed consent forms which were collated with their answer sheets to provide information such as age and gender for each subject. The time of day that each subject completed the test was also noted. This information was utilized to account for independent variables which may or may not have bearing on the dependent variable – the specific answers of each individual to each question.

To follow is an exploratory analysis of the data collected from the listening tests. Data are grouped in multiple ways in order to extract as many insights as possible. Data groups will be structured as follows:

- Responses for all respondents per question
- Responses for all questions per age-group
- Responses for all questions per time-of-day
- Responses for all questions per gender
- Responses per musical element implemented

1. Responses for all respondents per question

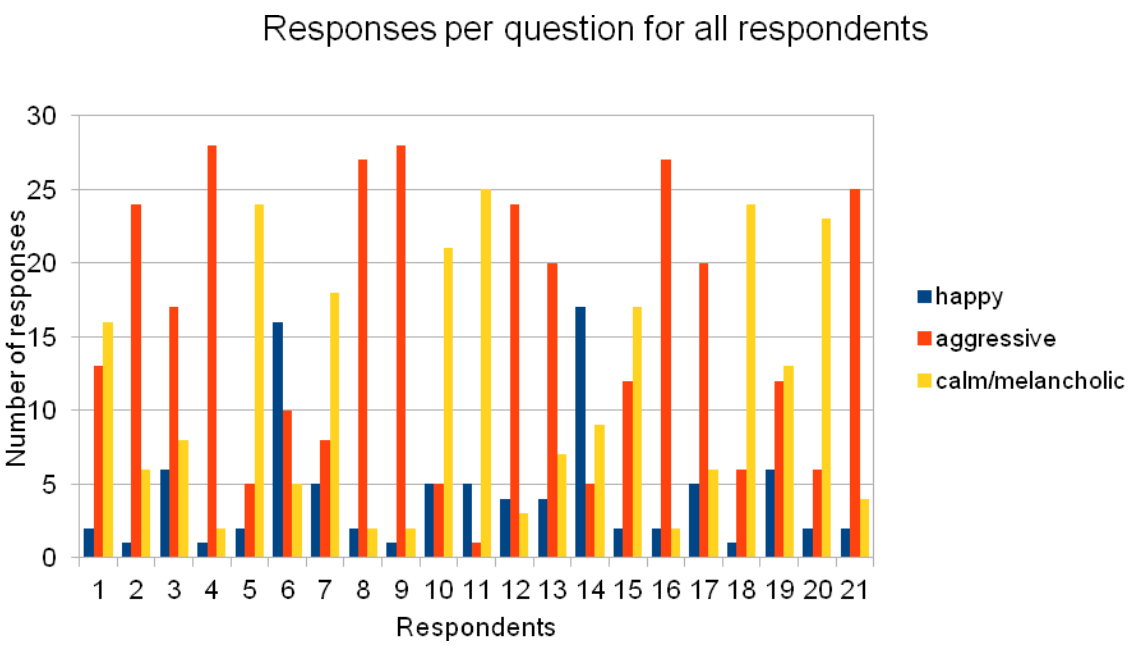


Figure 3: Responses per question for all respondents

The graph above shows the number of responses per question when all respondents' answers are combined (the individual graphs per respondent have been appended to this work so as to avoid unnecessary clutter here). What is interesting to note here is that the majority of the questions were ranked as either mostly aggressive or mostly calm/melancholic with only two out of the twenty one questions scoring as mostly happy. A significant indication, with regards to the aim of the study, is that on many questions there is a clear consensus in terms of response as to which descriptor was most commonly perceived in the listening material as many questions were answered with the same response by a majority of respondents.

Grand Totals:	**Total Answers**	**651**
	happy	91
	aggressive	323
	calm/melancholic	237
	% happy	14.0%
	% aggressive	49.6%
	% calm/melancholic	36.4%
	% high arousal	63.6%
	% low arousal	36.4%

Table 1: Totals for all responses and all respondents across happy, aggressive and calm/melancholic fields and for high vs. low arousal

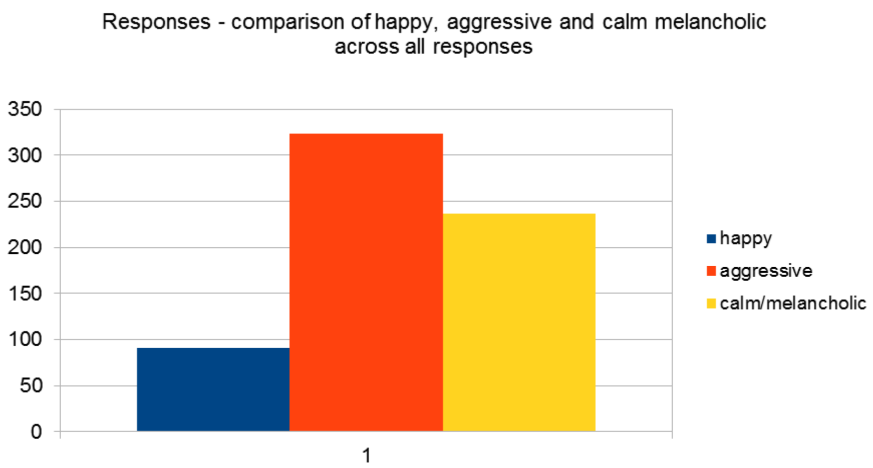


Figure 4: Responses – comparison of happy, aggressive and calm/melancholic across all responses

The totals for each emotion in terms of responses out of the possible 651 were:

- Happy: 91 (14.0%)
- Aggressive: 323 (49.6%)

- Calm/melancholic: 237 (36.4%)

When one considers the above in the light of Tolos et al. (2005) concluding that the arousal axis of Thayer's model is easier to define than the valence axis for most people, one might expect that the two high-arousal descriptors used in this test – happy and aggressive – might be more evenly balanced. Could one infer from this imbalance that the respondents in this study mostly perceive any high-arousal evoking white noise signal to be of a negative valence and thus select aggressive?

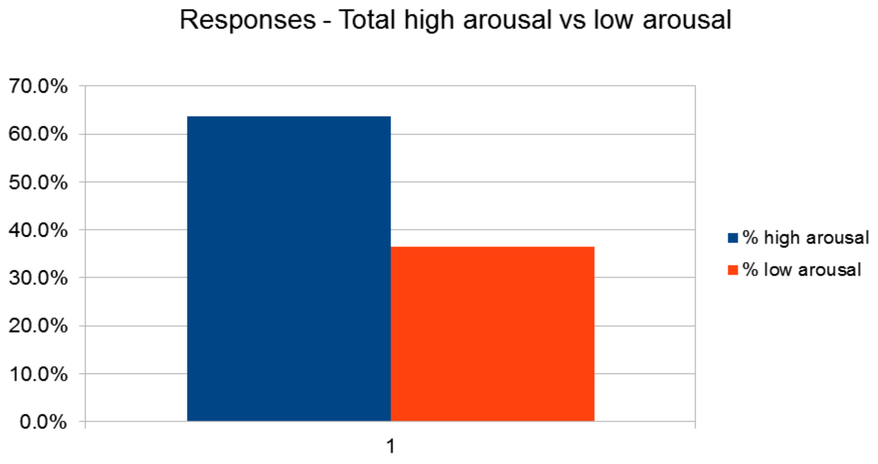


Figure 5: Responses – Total high arousal vs. low arousal

In order to give percentage answers per arousal-state, the data is combined by adding the answers for the valences of happy (positive) and aggressive (negative) to represent high arousal, and calm/melancholic remains as a representation of low arousal as there are already two valences represented here – calm (positive) and melancholic (negative). The data then shows that respondents found 63.6% of the listening material to have a high arousal and 36.4% to have a low arousal.

2. <u>Responses for all questions per age-group</u>

One of the independent variables which is given consideration here is the age of the respondents who took part in the listening tests. This seems appropriate as different age-groups of people do tend to appreciate different styles of music. There could also be bias against noise in a particular age-group which could skew results as they would score all the listening material negatively due to a predetermined negative association with noise-sound.

The ages of the 31 respondents were averaged and an average age of 39.61 years was the result. With this in mind I divided the respondents into 4 age groupings: below 30, 30-40, 41-50, and over

51. The results of the questionnaires were then calculated for each age-group, and can be seen in tabular and graphic form below.

Avg. Age		**39.61**			
	total responses	**no. of respondents**	**happy**	**aggressive**	**calm/melancholic**
Below 30	231	11	25	117	89
30 – 40	168	8	30	75	63
41 – 50	63	3	10	32	21
Over 51	189	9	26	99	64
	651		**91**	**323**	**237**
			happy %	**aggressive %**	**calm/melancholic %**
Below 30			10.8%	50.6%	38.5%
30 – 40			17.9%	44.6%	37.5%
41 – 50			15.9%	50.8%	33.3%
Over 51			13.8%	52.4%	33.9%

Table 2: Responses by age-group

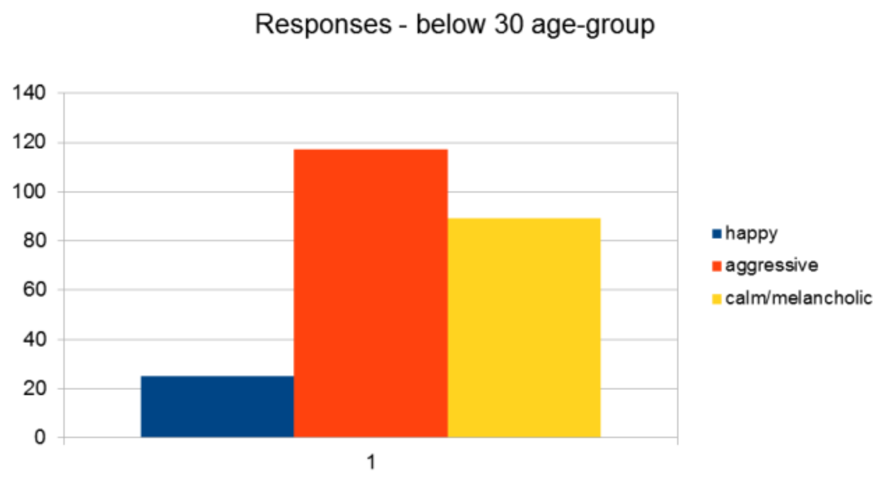


Figure 6: Responses – below 30 age-group

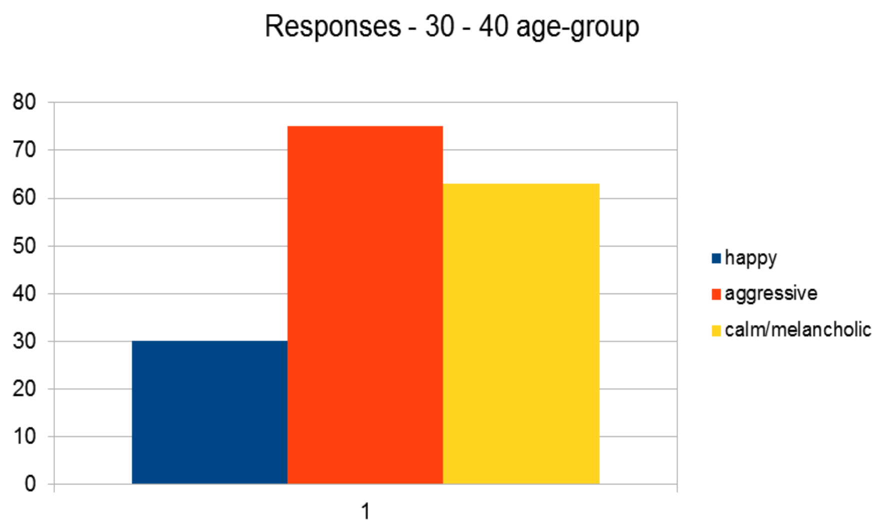


Figure 7: Responses – 30 – 40 age-group

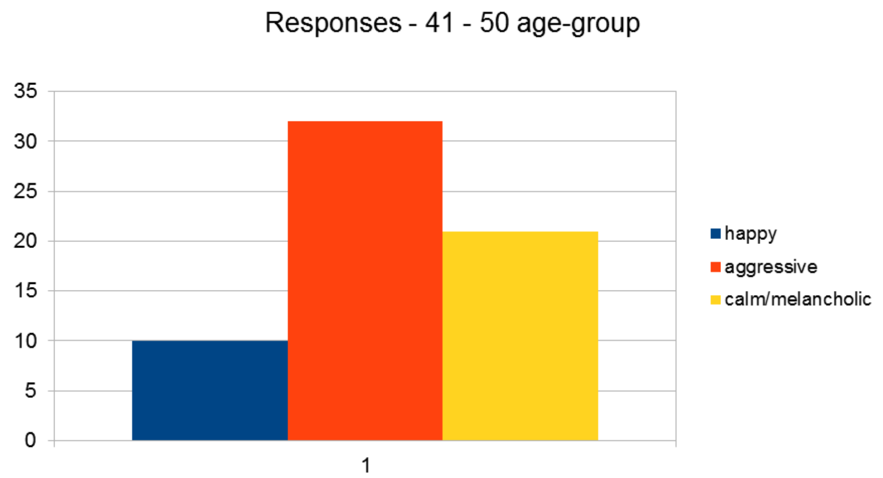


Figure 8: Responses – 41 – 50 age-group

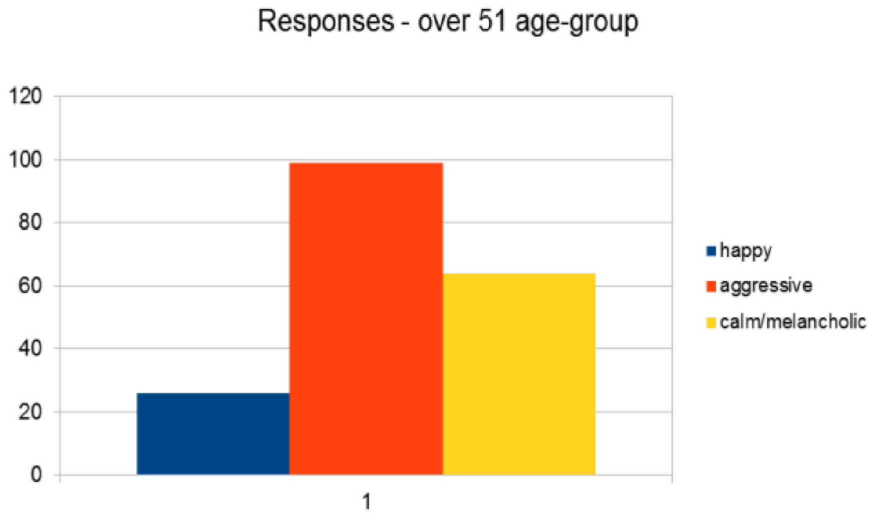


Figure 9: Responses – over 51 age-group

Of note here is how remarkably similar the responses for each of the age-groups are – 7.8% being the largest variance between any highest and lowest score. Also, almost counter-intuitively (as one might think they would be most amenable to a noise-signal being the youngest and most technologically minded), the under-30 age-group comes out with the least happy responses as they should be more comfortable with a medium such as noise due to its use in modern music and computer games. This is, however, only by 3% - not necessarily a major deviation. The similarity in responses must be cause for encouragement as it almost entirely eliminates age as a determinant of response.

3. <u>Responses for all questions per time-of-day</u>

Another independent variable which needed to be tested was the time-of-day at which each test was conducted. 'Morning' tests were all tests concluded before 12h00, and afternoon tests were any tests begun after 12h00.

Various factors related to the time-of-day could potentially influence the test-subject's responses. For example, morning tests may have been somewhat more exposed to background noise as the tests were conducted in an office at a school. Afternoon respondents may have suffered from fatigue caused by a stressful work day and thus not given the test their full concentration.

		happy	aggressive	calm/melancholic
Total tests morning	12	39	122	91
Total tests afternoon	19	52	201	146
Total	31	91	323	237
		%		
morning high arousal	161	63.9%		
morning low arousal	91	36.1%		
afternoon high arousal	253	63.4%		
afternoon low arousal	146	36.6%		

Table 3: Responses by time-of-day

To follow are graphic comparisons of morning and afternoon tests first by questionnaire response (happy, aggressive, calm/melancholic), then by arousal (happy plus aggressive vs. calm/melancholic), and finally a summary graph of arousal for morning and afternoon.

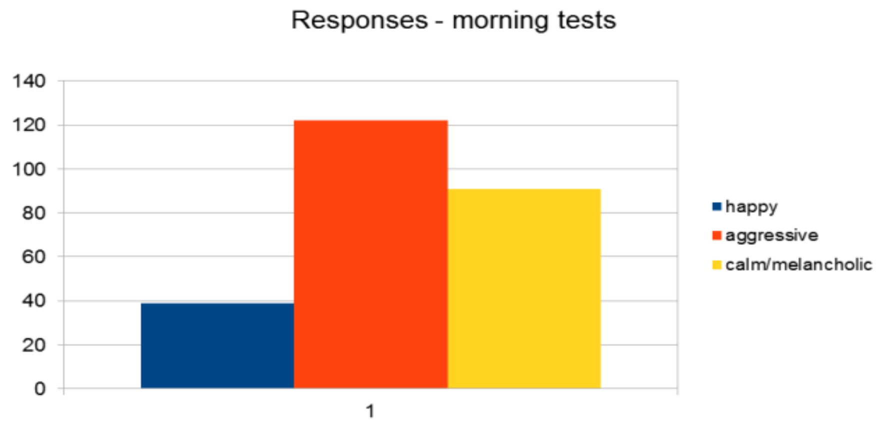


Figure 10: Responses – morning tests

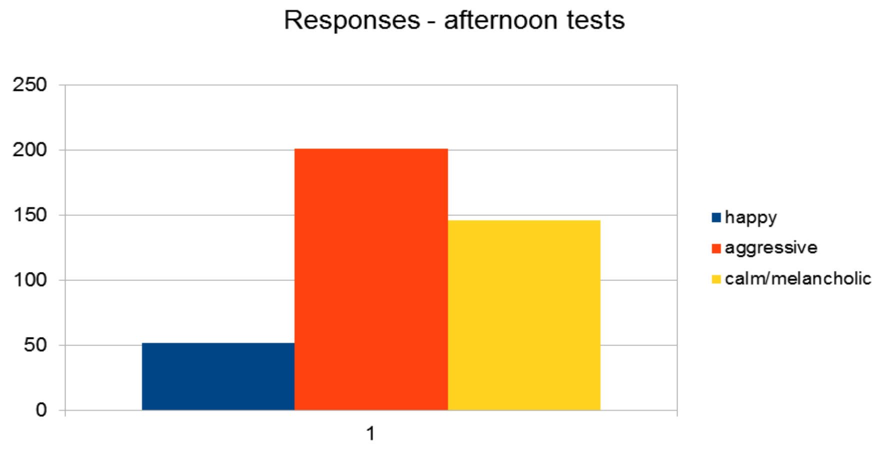


Figure 11: Responses – afternoon tests

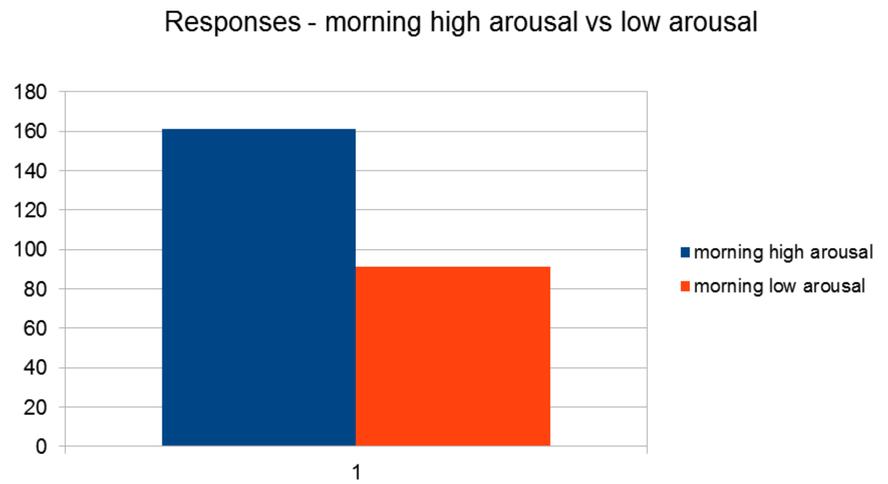


Figure 12: Responses – morning high arousal vs low arousal

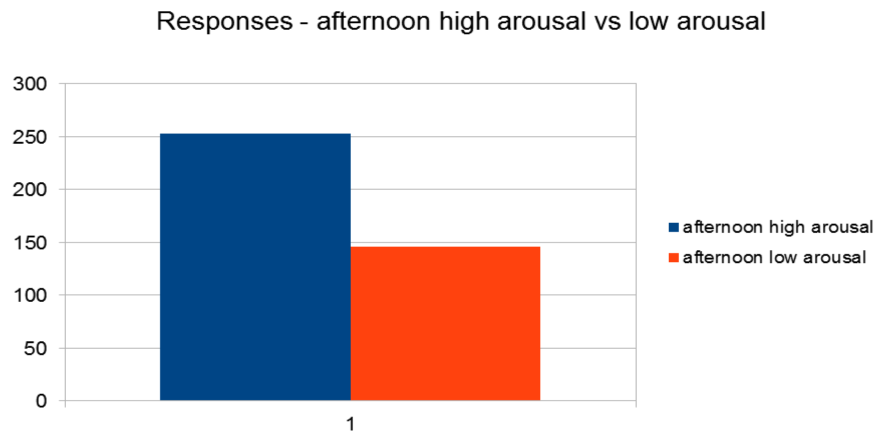


Figure 13: Responses – afternoon high arousal vs. low arousal

These illustrations all suggest that time of day was not an influential factor in the responses the test-subjects provided. Figure 12 below shows even more accurately how similar the scores are by arousal percentage from morning to afternoon.

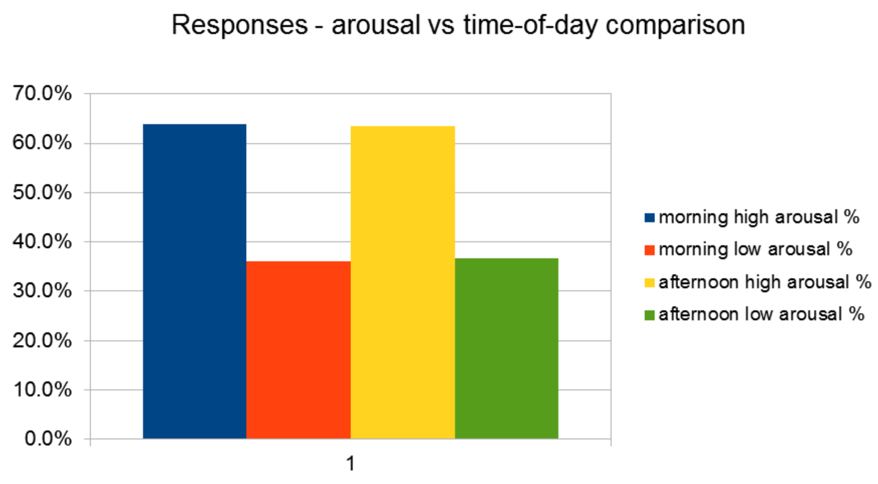


Figure 14: Responses – arousal vs time-of-day comparison

4. Responses for all questions by gender

The gender of each test-subject was another interesting independent variable which could be tested for its influence on the data results. Very simply, the responses of all male subjects were compared to the responses of all female subjects.

Male % happy	17.2%
Female % happy	11.6%
Male % aggressive	48.0%
Female % aggressive	50.8%
Male % calm/melancholic	34.8%
Female % calm/melancholic	37.6%

Male % high arousal	65.20%
Female % high arousal	62.43%
Male % low arousal	34.80%
Female % low arousal	37.57%

Table 4: Tabular summary of gender group responses 1

		happy	aggressive	calm/melancholic	
total males	13	47	131	95	273
total females	18	44	192	142	378

Table 5: Tabular summary of gender group responses 2

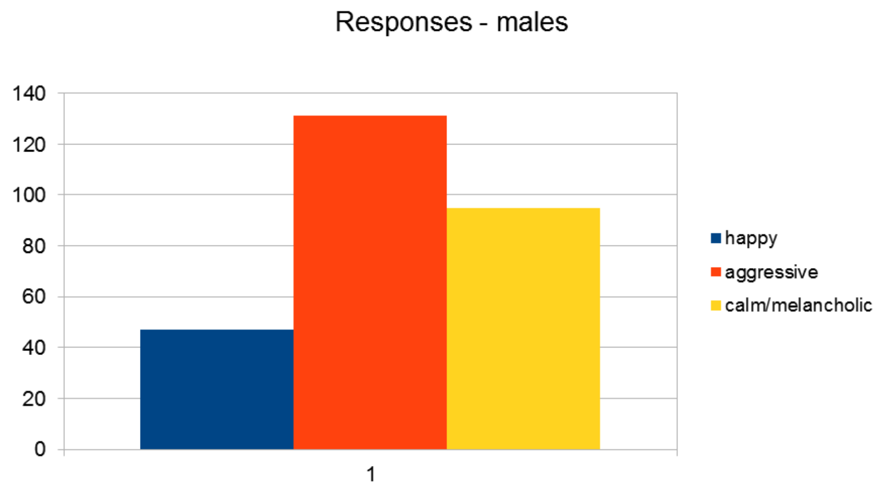


Figure 15: Responses - males

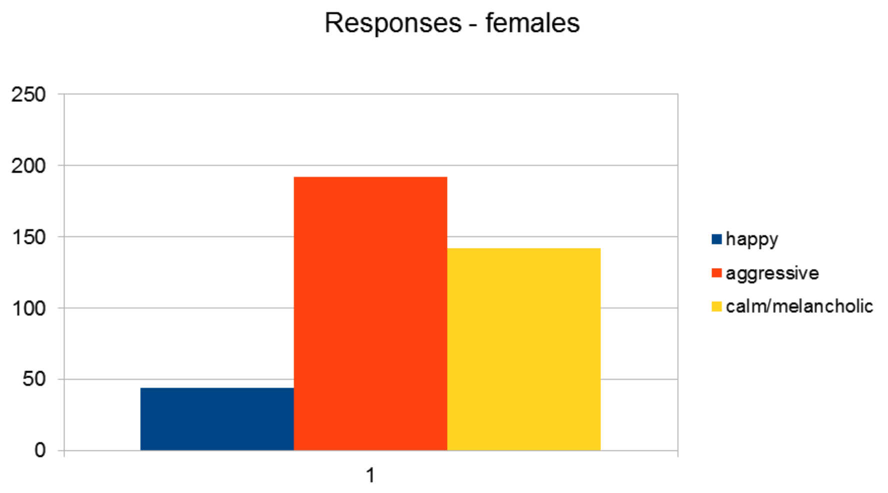


Figure 16: Responses - females

Although the above charts are markedly similar, the one noteworthy point is that men – once again contrary to perceived stereotype – responded with more 'happy' answers than did females (by 5.6%)

and the females with more 'aggressive' answers (2.8%). As such, in the following two charts, the men's results present with a higher value for high arousal than the women. This is, however, a marginal difference in the overall picture as it amounts to a mere 2.77% in terms of overall arousal level.

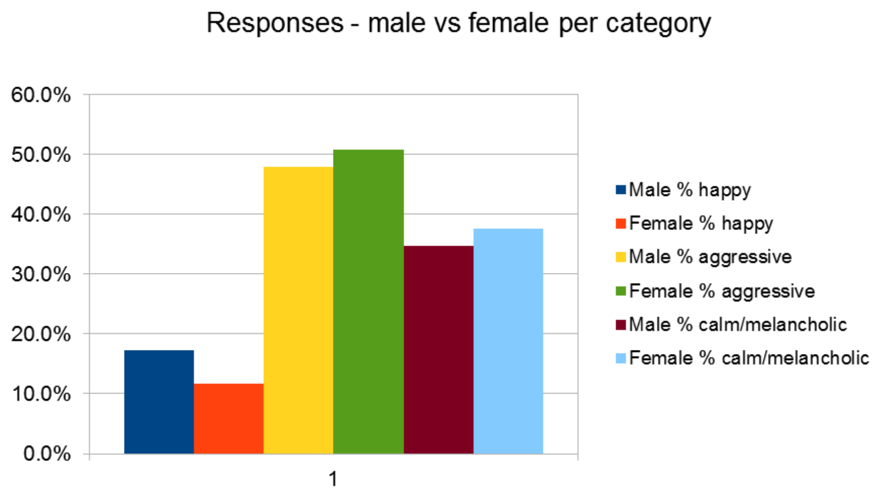


Figure 17: Responses – male vs. female per category

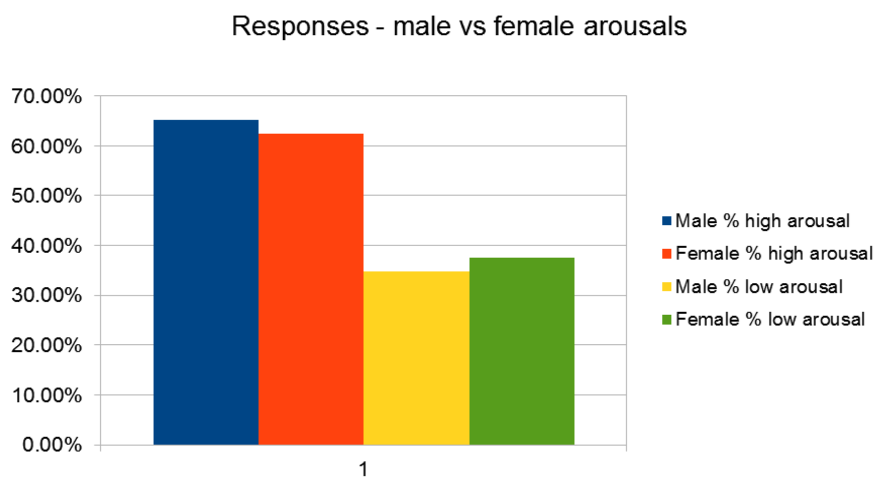


Figure 18: Responses – male vs. female arousals

Once again the above is a suggestion that, due to the nearly identical responses of men and women to the questionnaire, the gender of the individual respondent was not an influencing factor here.

With all of the above in mind the independent variables which were captured in the information sheets of the respondents have all been shown to have had very little, if any, impact on the responses to the listening test. The remaining grouping of data to be explored is the responses of all respondents to each musical element and DSP technique which was presented to them.

5. Responses per musical element and DSP technique combination

Musical elements and DSP techniques used were discussed in Chapter 5 and will not be thoroughly explained again here.

5.1. Dynamics

Questions 1, 9, 13 and 19 were the DSP noise-manipulations which mimic the musical element called 'dynamics'. Four dynamic effects were tested:

- small range; discrete change
- small range; continuous change
- large range; discrete change
- large range; continuous change

A 'small' dynamic change was regarded as any change within 6dB from nominal volume-level - 75dB, and a 'large' greater than 10dB from nominal volume-level. The test will consider both increase and decrease in volume from nominal level. The nature of the fluctuation of volume change was also tested. 'Abrupt' volume change was implemented through sudden volume change, whereas 'continuous' volume change employed a more gentle curve. These dynamic changes were implemented through the use of a white noise sample created in WAVELAB and played from the ReDrum unit in Reason. The volume was modulated by the automation of a volume control.

question	happy	aggressive	calm/melancholic
1	2	13	16
9	1	28	2
13	4	20	7
19	6	12	13

Table 6: Responses per category per question - dynamics

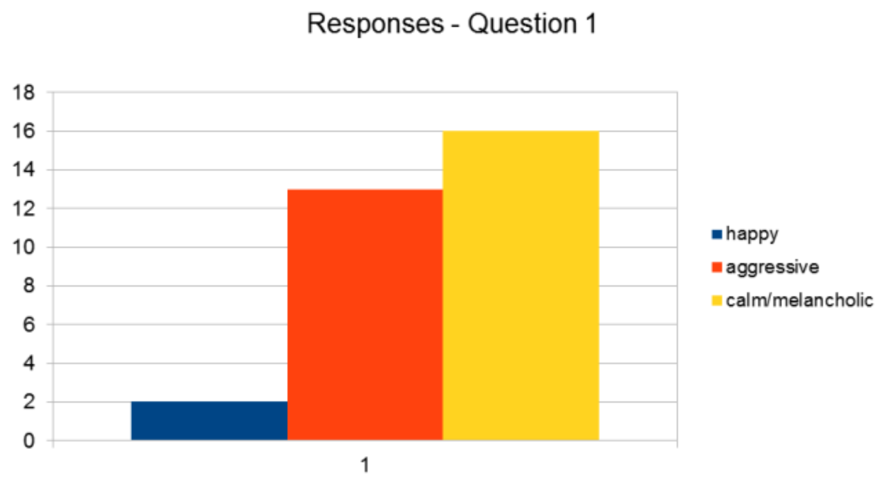


Figure 19: Responses – Question 1

Question 1 measured responses to a dynamic fluctuation with a small volume range and continuous volume change. The responses are represented graphically below. Responses here were split between calm/melancholic (16) and aggressive (13) with only 2 happy responses. This kind of close split could indicate that little musical effect was clear to respondents in this material.

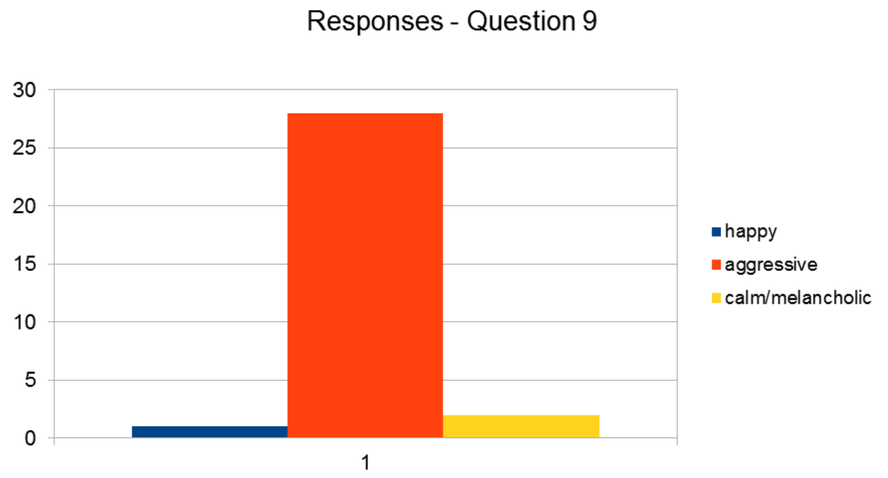


Figure 20: Responses – Question 9

Question 9 was a simulation of a dynamic fluctuation with a large volume range and abrupt volume changes. The responses to this test were overwhelmingly aggressive – 28 out of 31. The graph above illustrates this. This is a clear indication that the musical effect for this type of dynamic flux was one of high arousal and low valence.

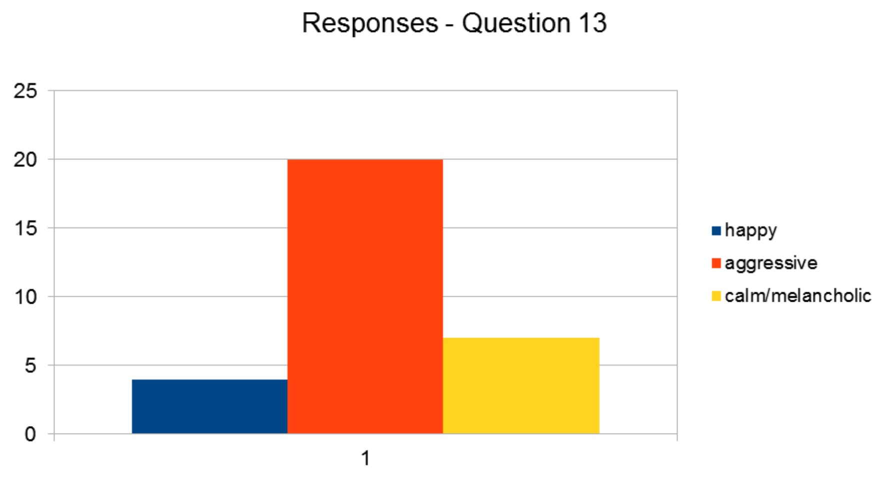


Figure 21: Responses – Question 13

The material on question 13 was a dynamic fluctuation with a large volume range and continuous level change. The responses here were again mostly aggressive (20 out of 31), however, somewhat less so than for question 9. When one compares these two, the differentiating factor is the type of volume level change presented. In question 9 the volume changes were abrupt, whereas in question 13 they were continuous. This continuity could account for more respondents feeling more at ease with the material in question 13.

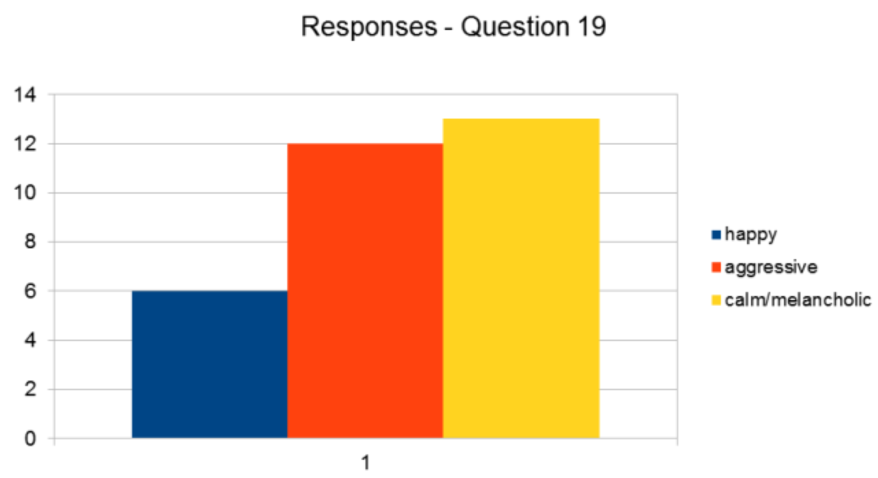


Figure 22: Responses – Question 19

The responses to question 19 – a dynamic fluctuation with a small volume range and abrupt volume changes are similarly spread to those in question 1. That both examples of small dynamic range were perceived as ambiguous is not necessarily problematic due to information discussed in Chapter 3 which will be revisited in the discussion section of this treatise.

5.2. Spectral Shape

Questions 2, 7, 15 and 20 were the DSP noise-manipulations which mimic the musical element called 'spectral shape'. Alteration of spectral shape was conducted through the use of a white noise sample created in WAVELAB and played from the ReDrum unit in Reason. The noise was manipulated by the control of a shelving filter.

Four variants of spectral shape were included in the test. 'Bright' refers to high frequency energy above 2kHZ being enhanced by 10dB. 'Dark' refers to low frequency energy below 200HZ being enhanced by 10dB. 'Mid-heavy' refers to mid-frequency energy from 200 Hz to 2kHz being enhanced by 10dB. 'Mid-lite' refers to mid-frequency energy from 200Hz to 2kHz being decreased by 10dB.

Spectral shape effects tested:

- bright
- dark
- mid-heavy
- mid-lite

question	happy	aggressive	calm/melancholic
2	1	24	6
7	5	8	18
15	2	12	17
20	2	6	23

Table 7: Responses per category per question – spectral shape

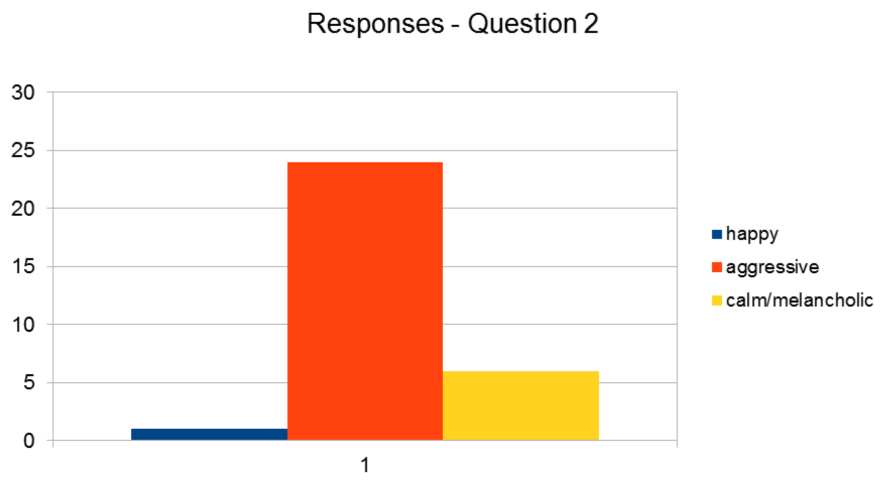


Figure 23: Responses – Question 2

Question 2 presented a 'dark' spectral shape to the respondent. The vast majority of the feedback from this question indicates that this is an aggressive sound with 24 out of 31 respondents marking it as aggressive.

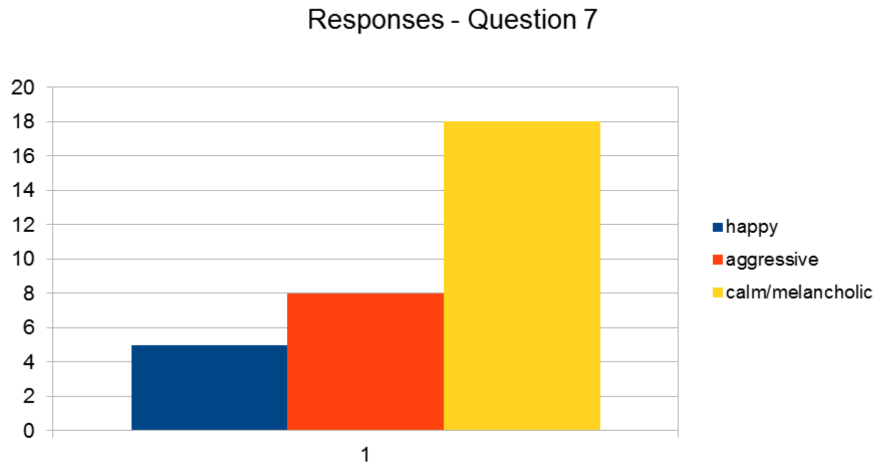


Figure 24: Responses – Question 7

Question 7 presented a 'mid-heavy' spectral shape, which 18 out of 31 respondents found calm/melancholic in nature. 8 perceived it as aggressive and 5 as happy. This result appears reasonably conclusive in favour of calm/melancholic until one considers the high versus low arousal scores – 18 low to 13 high. This suggests that a substantial number of respondents could not agree on whether this was a high or low arousal effect.

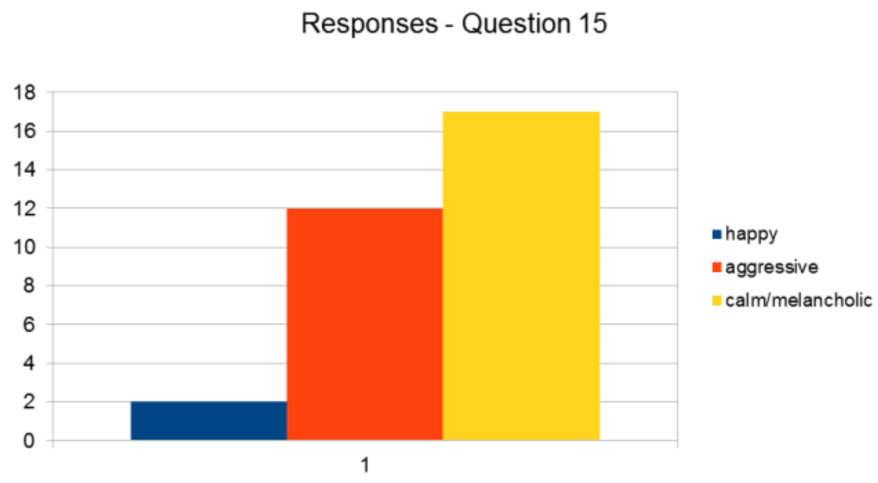


Figure 25: Responses – Question 15

A 'bright' spectral shape was produced for question 15. Here the results were split between calm/melancholic with 17 and aggressive with 12. Only 2 respondents chose 'happy' in response.

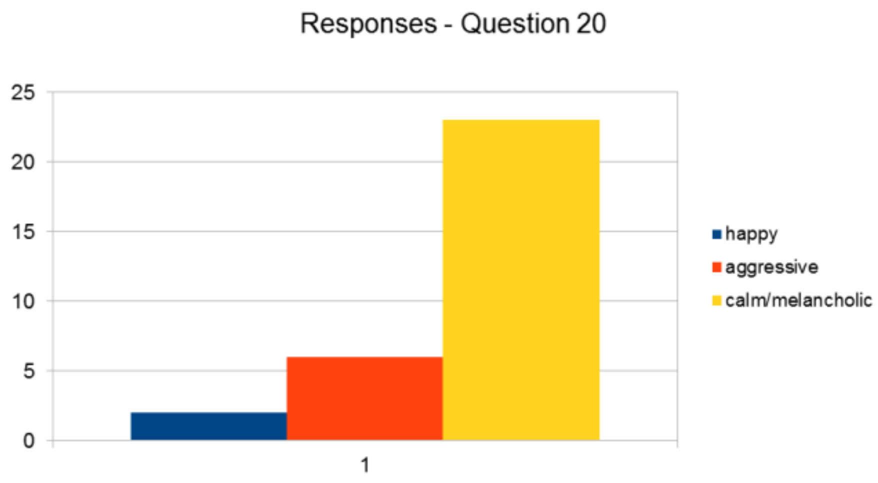


Figure 26: Responses – Question 20

The audio in question 20 was made to present a noise signal which was 'mid-lite' in spectral shape. This question was perceived as calm/melancholic by 23 of the 31 subjects in the test.

The data collected for the questions in which spectral shape was emulated are interesting as a dark spectral shape is the only example from the four variants tested which evoked a high arousal-state (aggressive) in respondents. The other three – bright, mid-heavy and mid-lite all garnered an opposite response (low arousal-state). The one thing which does present itself is that the aggression rating is highest at opposite ends of the testing material, that is, the dark spectral shape and bright spectral shape received the most 'aggressive' ratings (24 and 12 respectively), whilst the two dealing with mid frequencies were scored predominantly toward calm/melancholic. This phenomenon leads to a couple of questions: do people perceive the mid frequency exaggerations as more calm/melancholic because mid frequencies are in the band within which they hear most easily – making them more comfortable, and conversely perceiving frequencies on the outer-limits of their hearing as aggressive as a high arousal is generated through more audio activity in the extreme edges of our hearing spectrum?

5.3. Rhythm

Questions 3, 6 and 14 were the DSP noise-manipulations which mimic the musical element called 'rhythm'. Rhythms are differentiated through their complexity values, a concept first researched by Lerdahl and Jackendoff in their authoritative text 'A Generative Theory of Tonal Music' (1983). Toussaint expanded on their work with his Metrical Complexity Measure proposed in his texts of 2002 and 2007 wherein he calculates a sum of the metric weights for each onset in a given rhythm and assigns a complexity value based on that sum – the lower the value the more complex the rhythm.

3 Rhythm patterns were tested:

- Simple
- Medium
- Complex

Toussaint's formula for rhythmic complexity assigns a numeric value to each pulse in a given bar. A 4/4 bar is said to have 4 beats and 16 pulses. Below is a graphic representation of the values he assigned to each pulse in a bar.

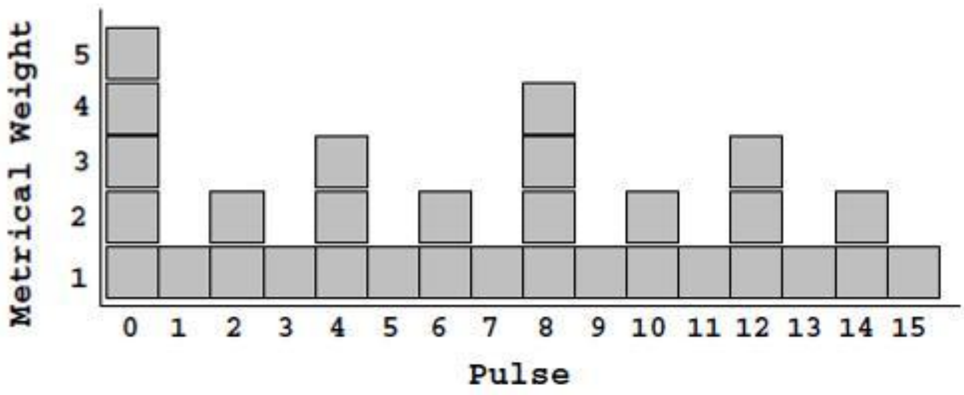


Figure 27: Toussaint's Metrical Complexity Measure for a bar of 4 beats

In order to test the evocative effects of rhythm 3 rhythms of varying complexity were generated. A 'simple' rhythm with a complexity measure of 25, a 'medium' rhythm with a complexity measure of 17, and a 'complex' rhythm with a complexity measure of 10.

Rhythms were generated through the use of a white noise sample created in WAVELAB and played from the ReDrum unit in Reason. The noise was manipulated by the control of a bandpass filter, which creates the sensation of pattern without volume alteration.

question	**happy**	**aggressive**	**calm/melancholic**
3	6	17	8
6	16	10	5
14	17	5	9

Table 8: Responses per category per question - rhythm

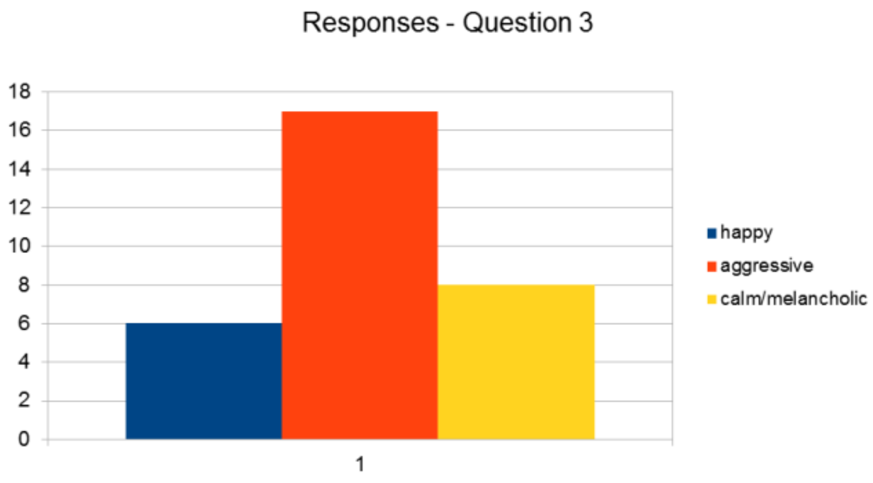


Figure 28: Responses – Question 3

The rhythm in question 3 was of a medium complexity. As can be seen in the graph above, 17 respondents scored this as an aggressive sound, 8 as calm/melancholic and 6 as happy. This means that 23 of 31 respondents thought a high arousal effect was present.

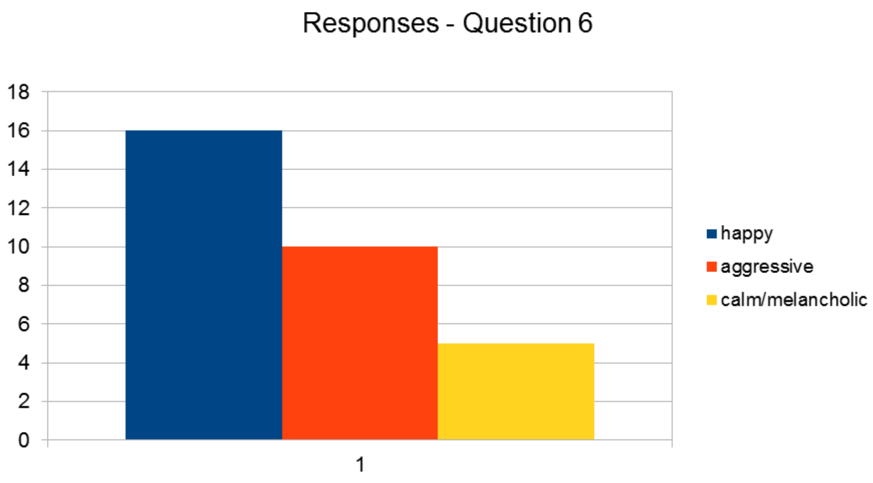


Figure 29: Responses – Question 6

Question 6 was a rhythm with a simple complexity score. 16 respondents rated this question as happy, 10 as aggressive and 5 as calm/melancholic (see graph above). This result gives the signal a heavily weighted high arousal score at 26 out of 31 responses.

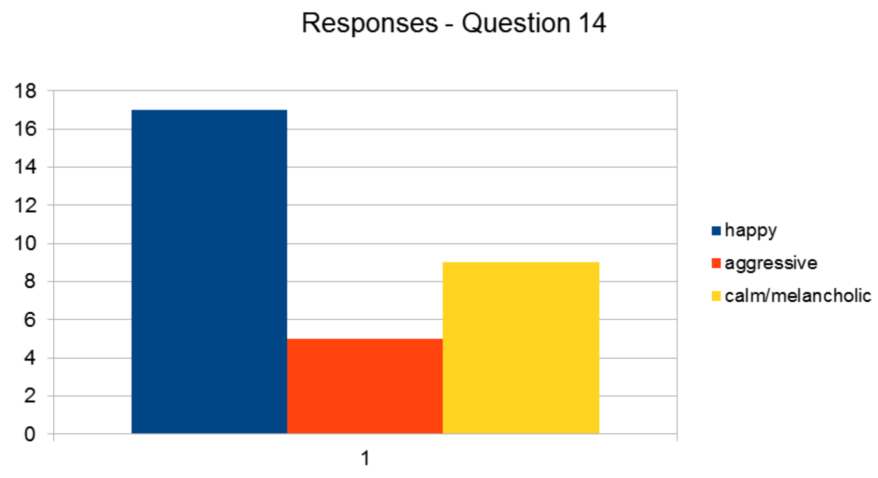


Figure 30: Responses – Question 14

Question 14 represented a rhythm with a high degree of complexity. As per the graph above, 17 respondents rated this as happy, 5 as aggressive and 9 as calm/melancholic. Once again, the majority (22) found this to be a high arousal evoking sound.

What is interesting with regard to the responses for the questions displaying various rhythmic complexities is that it would be expected that if the simplest rhythm was the happiest then the medium one should be second happiest and the complex rhythm should be the least happy. This, however, has not proved to be the case in the current research. The most complex rhythm rates as the happiest, the simple one second and the medium-complexity example third. The arousal scores confirm this as, once again, the simplest rhythm should rate as lowest arousal and the most complex as highest arousal, and yet this order is reversed. Another surprise here is that the medium-complexity rhythm was mostly perceived as aggressive whereas both the simple and complex ones were perceived mostly as happy. This is a somewhat unexpected result. If the data showed that one of the examples on the outer edge of the example set was perceived as markedly different from the others that would be more understandable, but to have the middle example score opposite to the two outer ones in the set is difficult to grasp. This may warrant further investigation.

5.4. Peak Sound Level

Questions 4, 8, 12 and 17 were the DSP noise-manipulations which mimic the musical element called 'peak sound level'. The test of the evocative nature of peak sound level was administered by the increase in sound level of a portion of the noise signal by 30dB (from 65dB SPL to 95dB SPL) and then a decrease back to 65dB SPL.

Four variants of this process were tested in order to gain understanding of whether or not time had a part to play in the effect created by peak sound level perception. A slow increase – decrease over 15 seconds, a medium increase – decrease over 10 seconds, a more rapid increase – decrease over 5 seconds, and a sudden increase – decrease over 1 second were all implemented in the test to provide a variety of values with regard to the time factor.

4 Peak sound level tests:

- 30dB slow (15 seconds)
- 30dB medium (10 seconds)
- 30dB fast (5 seconds)
- 30dB sudden (1 second)

These effects were generated through the use of a white noise sample created in WAVELAB and played from the ReDrum unit in Reason. This noise was modulated in volume through the use of a volume control whose automation was timed to achieve the two desired cycles.

question	**happy**	**aggressive**	**calm/melancholic**
4	1	28	2
8	2	27	2
12	4	24	3
17	5	20	6

Table 9: Responses per category per question – peak sound level

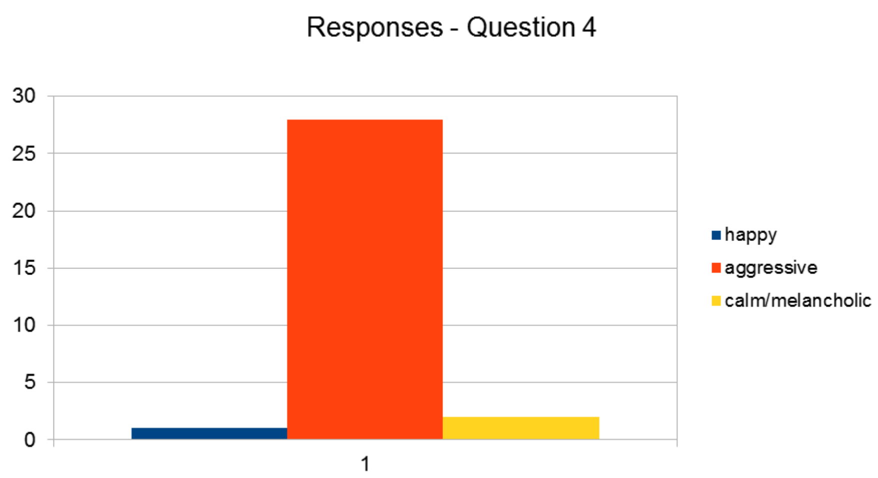


Figure 31: Responses – Question 4

Question 4 was a demonstration of a peak sound level achieved at what was, for this test, referred to as a medium duration. 28 out of 31 respondents rated this as an aggressive effect, 2 as calm/melancholic and only 1 as happy as seen in the graph above.

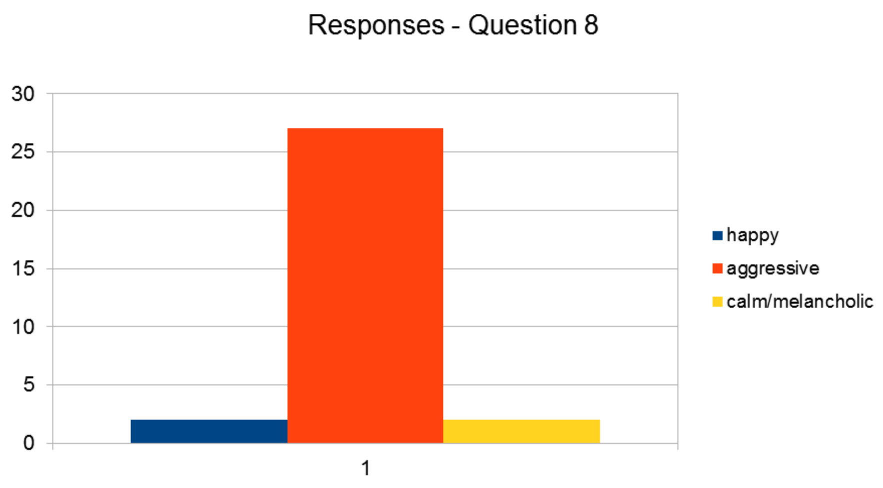


Figure 32: Responses – Question 8

A 'sudden' peak sound level was tested in question 8. Aggressive received 27 of the 31 possible responses with happy and calm/melancholic each getting 2 of the remaining 4. This is shown in the graphic representation above.

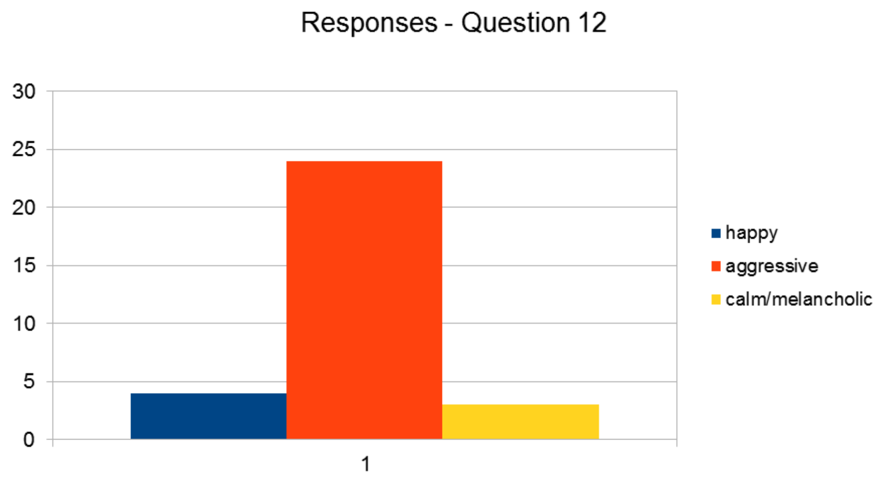


Figure 33: Responses – Question 12

The audio material in question 12 was acoustic representation of a slow sound level peak. As per the chart above 24 respondents thought it was aggressive, 4 happy and 3 calm/melancholic.

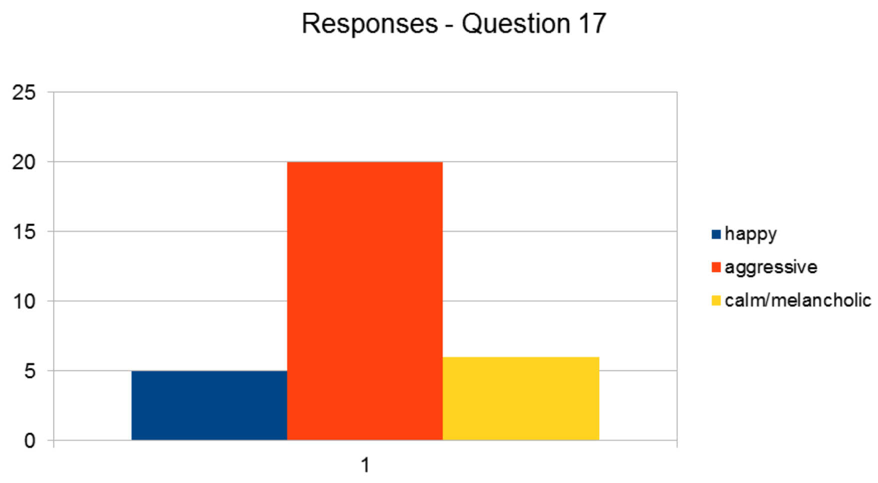


Figure 34: Responses – Question 17

In question 17 the peak sound level was achieved in a short space of time – called 'fast' for the purposes of this study. In the graph to follow one can see that 20 subjects perceived this as aggressive, 5 as happy and 6 as calm/melancholic.

The results of the peak sound level tests suggest that a volume change as large as 30dB, whether it occurs suddenly or more slowly, will evoke a high arousal – negative valence state. This is evidenced by the fact that each of the four audio examples was identical apart from the rate at which the volume change occurred, and each of the four time variants tested received an overwhelming majority of 'aggressive' responses. The only question one could ask here is how would that effect be different if the signal being transmitted was not noise, in other words, how much does the noise have to do with the arousal and valence attached to the perceptions of the respondents?

5.5. Repetition

Questions 5, 10, 11, 16, 18 and 21 were the DSP noise-manipulations which mimic the musical element called 'repetition'.

6 Repetition Tests:

- 125ms
- 250ms
- 500ms
- 750ms

- 1 second
- 2 seconds

For this part of the experiment a short (2 second) segment of our original white noise signal was cut out and then repeated 16 times to create a continuous 32 seconds of noise. This effect was created in WAVELAB as REASON was not needed.

question	happy	aggressive	calm/melancholic
5	2	5	24
10	5	5	21
11	5	1	25
16	2	27	2
18	1	6	24
21	2	25	4

Table 10: Responses per category per question - repetition

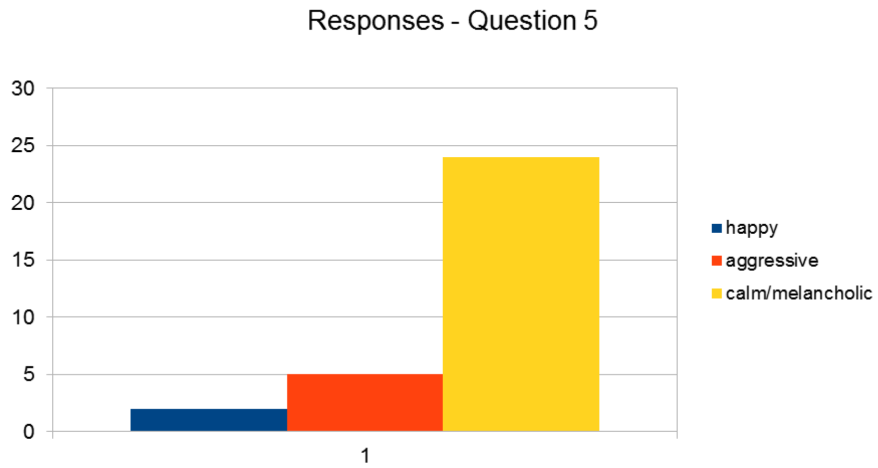


Figure 35: Responses – Question 5

Question 5 was the longest repetition (a two second segment) of the 6 tested. 24 of the 31 respondents perceived this as calm/melancholic, 5 as aggressive and only 2 as happy.

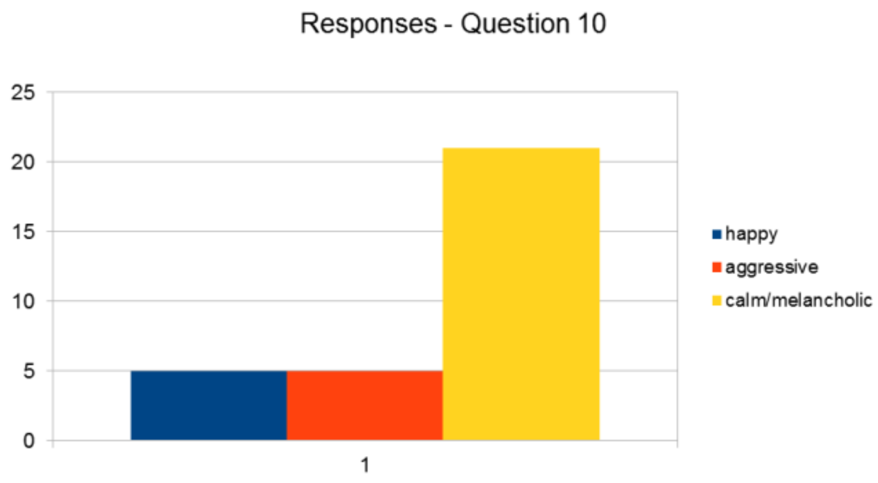


Figure 36: Responses – Question 10

Question 10 consisted of a 500ms segment of noise which was repeated multiple times. 21 subjects rated this as calm/melancholic with the other 10 responses evenly split between happy and aggressive.

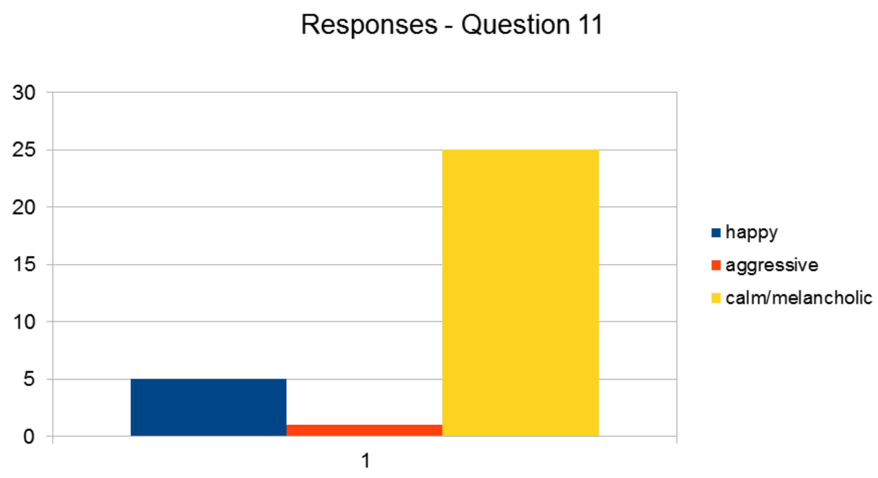


Figure 37: Responses – Question 11

Responses to the 1 second repetition in question 11 were heavily weighted toward calm/melancholic which received 25 of the 31 possible responses. Happy was chosen 5 times and aggressive only once.

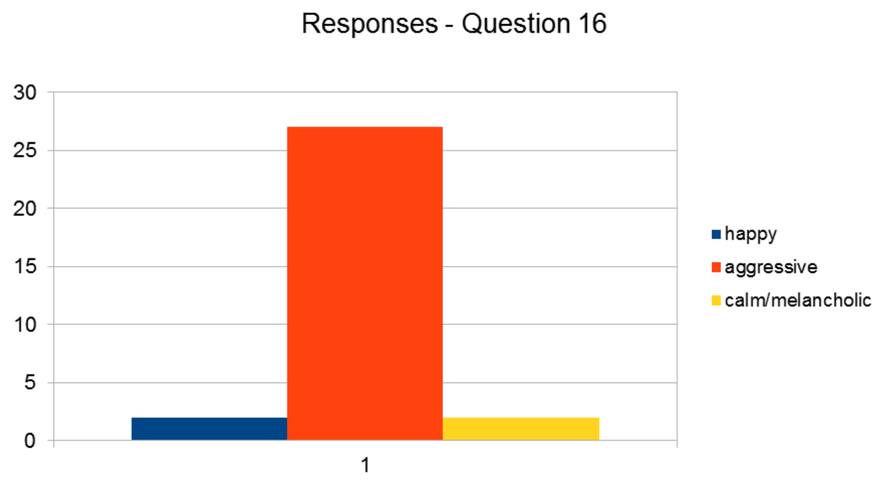


Figure 38: Responses – Question 16

Question 16 was a very short repetition – 125ms. 27 subjects chose aggressive in response to this material. The other 4 responses were split between happy and calm/melancholic.

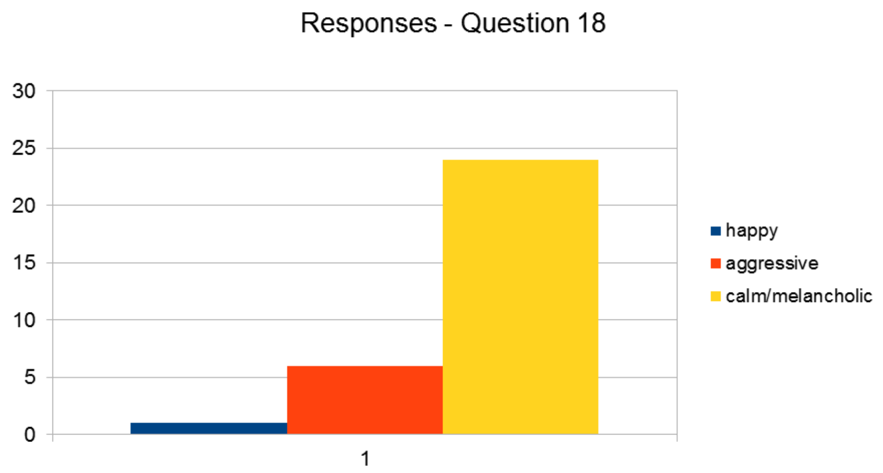


Figure 39: Responses – Question 18

A 750ms repetition was created for question 18. of the 31 respondents 24 perceived this as calm/melancholic, 6 as aggressive and 1 as happy.

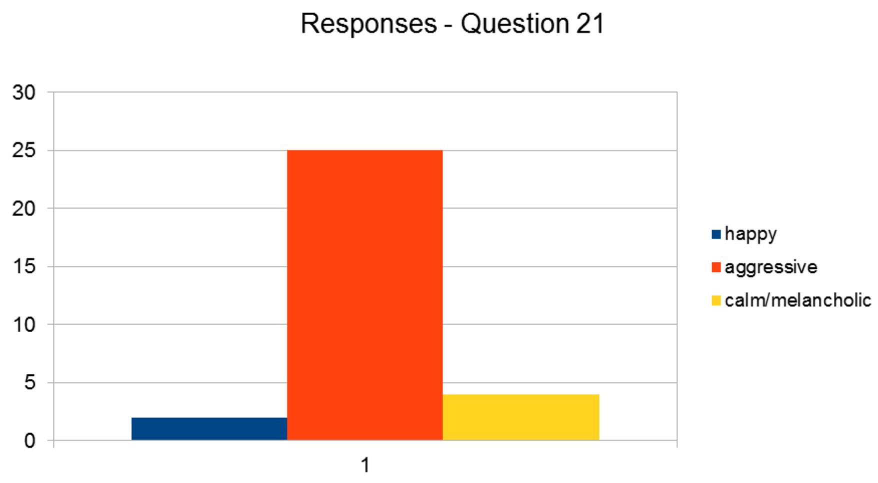


Figure 40: Responses – Question 21

The last question in the listening test, question 21, was a 250ms repetition. 25 of 31 test-subjects rated this as aggressive, 4 as calm/melancholic and 2 as happy.

Two main factors emerge as being noteworthy from the tests of repetition: firstly, the human brain is clearly able to discern repetition of a random signal – the material in the test was a random piece of specified length cut out of a noise signal and repeated. The fact that the data changed through the series of repetition tests shows that it had an effect and as such was recognised by the brain. Secondly, any repetition of 500ms or more was perceived as largely calm/melancholic and anything of 250ms or shorter was overwhelmingly perceived as aggressive.

6.2 CHAPTER SUMMARY

Chapter 6, the data analysis chapter of the current research, set-about the task of tabulating, presenting, exploring and describing the features of the raw data collected at the listening experiments. The data was examined across numerous criteria in order to better understand the responses given by the test-subjects. Various independent variables were considered, and interesting points of observation highlighted in order to prepare the way for a more thorough discussion of these results in Chapter 7.

CHAPTER 7 - CONCLUSIONS

7.1 DISCUSSION OF RESULTS

The results of the listening experiments presented in Chapter 6 provide a framework for this final chapter that discusses the findings of the research in terms of the research aims and brings the present study together in a conclusive summary. The chapter summarizes and discusses the key research findings within the theoretical and research contexts proposed in earlier chapters. In addition, the chapter acknowledges the limitations of the present study and provides recommendations that may assist future researchers with similar studies.

This quantitative study employed a two-part design, first focussing on an investigation of the literature in each of the relevant subject-areas, and then implementing the knowledge gained from this investigation in the form of a listening experiment with the goal of testing the specific aims of the research. These aims, summarized in the form of four questions, are as follows (question 1 being the primary research question):

1. Is it possible to evoke the musical effect using real-time noise control methods?
2. Which musical elements are the most evocative of the musical effect when implemented via a noise-control technique?
3. Which properties of digital signal processing methods are best suited for use in the creation of noise signals which evoke the musical effect?
4. What are the possible applications of these techniques in both the music world, and other spheres?

In order to begin to answer these questions a literature study was conducted. This study was divided into two parts as the research question falls into two distinct bodies of knowledge – music and technology. In Chapter 3 the concept of music and the human experience thereof was discussed. Music was found to be best defined as 'organised sound which causes an emotional response in the listener'. Music was then disassembled into its component parts so as to make the experimental phase of this research more manageable. Each of these elements of music was investigated for its affective strengths within the musical whole. The fields of research called Music Emotion Recognition and Music Emotion Synthesis were also explored, and here knowledge was gained relating to emotion identification in music and the tools which are being used to synthesize emotion in computer-generated music. The other area involved in the literature scan, Digital Signal Processing, was described in Chapter 4 of this treatise. First, a foundation was laid through an investigation of the basic principles involved in DSP and noise was conceptualised and its variants discussed. Thereafter, common DSP operations, digital sound effects and synthesis were explored with the hope of uncovering some possible techniques which could be used to manipulate noise in the listening experiments.

The culmination of this two-pronged approach to the literature review was the design of a listening experiment to be completed by at least 30 adult respondents – 31 were actually tested – from Port Elizabeth, South Africa. The experiment material was created in real-time in Studio 2 at the Nelson Mandela Metropolitan University prior to the experiments. This was to ensure that there was no possibility of error in the reproduction of these sounds from respondent to respondent in the experiments. The experiment was conducted over the course of two days in a controlled environment on high-quality audio reproduction equipment. Respondents listened to the test material whilst simultaneously answering the questions on the questionnaires designed for the tests. The experiment tested 21 variants of digitally manipulated white noise and respondents answered either 'happy', 'aggressive' or 'calm/melancholic', based on their perception of the emotional quality of each sound. These emotion labels were found in Tolos et al. (2005) to be universally sound as labels to be used in experiments such as this one for categorising emotion according to the Thayer model of affect. Once the experiments were completed, the data from the questionnaires was tabulated and Chapter 6 of this treatise shows various graphic representations and analyses of this data.

To follow, is a look at each of the questions previously stated as the aims of the current research, in order to attempt to provide possible answers based on the data from the experiments and the knowledge gained from the literature review.

7.1.1 Is it possible to evoke the musical effect using real-time noise control methods?

The main aim of the current research has been to ascertain whether or not it is possible to evoke the 'musical effect' – or the emotional response people experience when they listen to music – using white noise manipulated via real-time digital techniques. In order for the answer to this question to be 'yes', there should be a clear indication in the responses to the listening experiments that the respondents experienced some emotional affect in the sounds they were exposed to. As they were forced to choose a particular option in response to each question, the only conclusive evidence of this would be if there are instances in the data where a large proportion of respondents selected the same option, or at worst, an option from the same arousal domain (because, as stated above, valence is more difficult for subjects to identify than is arousal). As can be seen quite clearly in Figure 1 (reproduced from Chapter 6) below, there are a number of such points of agreement amongst respondents.

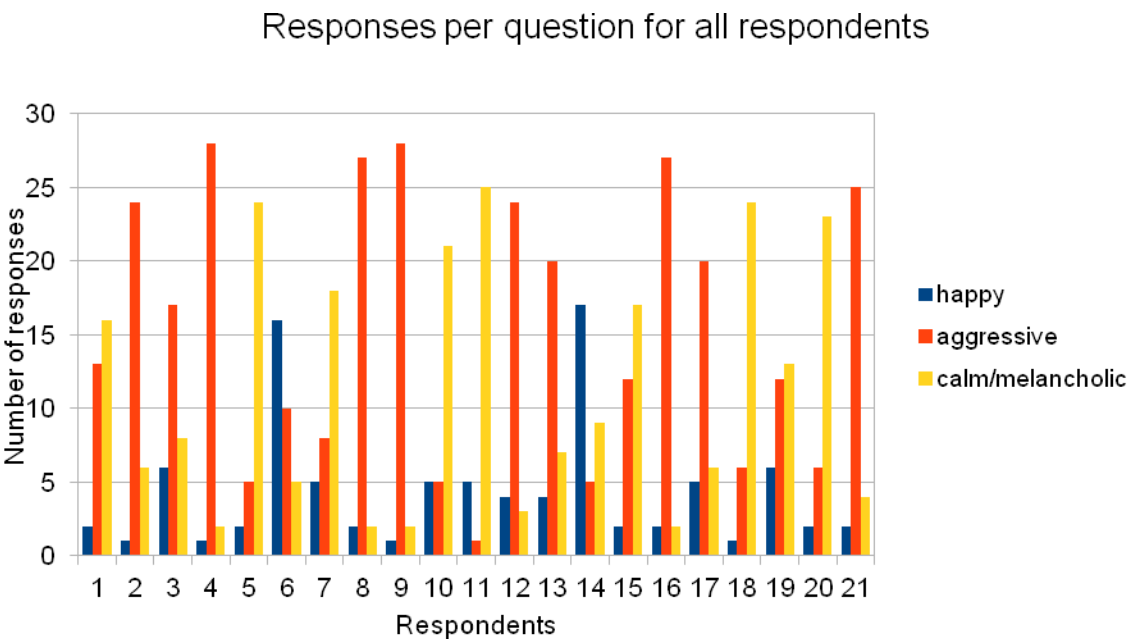


Figure 41: Responses per question for all respondents

Of course, this does not prove exactly how strongly the emotion was felt, and cannot be taken as totally conclusive, as some of the responses to any of these questions could include responses from subjects who were guessing or simply choosing randomly. However, the chance is very small that there could be this many questions to which such a significant number of respondents all guessed the same random response. As such, one can conclude that an emotional response was cause of the choice made by many of the respondents – a conclusion which leads me to say that the answer to the research question is 'Yes, it is possible to evoke the musical effect through the manipulation of noise via real-time digital techniques'. Of course, there are some limiting factors which must also be considered. The responses cannot be generalised to the population outside of Port Elizabeth, South Africa. Nor can they be viewed as a reflection for all age-groups, as only adults over the age of twenty one were tested. The results reflected in the data analysis chapter do however lead to the conclusion that gender, time-of-day and age (over twenty one) do not affect perception with regard to the emotion perceived in the experiment.

7.1.2 Which musical elements are the most evocative of the musical effect when implemented via a noise-control technique?

As discussed in Chapter 5, five elements of music emerged from the literature review process as being the most likely to be replicable when a noise signal is manipulated via DSP techniques. These were:

1. Dynamics
2. Spectral Shape
3. Rhythm
4. Peak Sound Level
5. Repetition

In order to make inferences as to which elements were the most evocative when implemented via a noise control technique, one must identify the elements that were most consistently rated as evocative in the same emotion or arousal rating by respondents across all tests conducted for that element. Results for each element tested will be summarized below to gain insight into the success rate of each element.

7.1.2.1 Dynamics

In Chapter 3 evidence was uncovered to suggest that composers use dynamics to shift the focus of the listener (Meyer 1956). Of the four implementations of dynamics the two examples which represented large dynamic range were conclusively rated as high arousal effects. However results for the two examples of small dynamic range were inconclusive. If the finding from Chapter 3 is taken into account, then one could say that the large dynamic range examples achieved their objective, and that, although the responses to small dynamic range were ambiguous, the possibility exists that small dynamic range could be ambiguous in meaning all the time (and thus validate the findings of the experiment) since music with small dynamic range may not be intended to shift the listener's focus at all. This question falls outside of the scope of the current research, however it could make for interesting further study in another text.

7.1.2.2 Spectral Shape

This was possibly the least successful element tested as only 'dark' and 'mid-lite' spectral shapes garnered any substantially similar response. 'Dark' spectral shape was classified as aggressive by an overwhelming majority of subjects whilst 'mid-lite' was classified as calm/melancholic. These results are both in agreement with information captured in Chapter 3 regarding timbre – the major contributor to spectral shape. The information in Chapter 3 may also account for the two seemingly less successful elements tested, as there is evidence that perception of timbre is situational. If this is true then, perhaps, the respondents were unable to decide the emotion conveyed by these two spectral shapes due to the unfamiliar nature of the noise they were being exposed to. Whether or not this is the case, the data does suggest that spectral shape is not a particularly effective musical element when recreated using noise as a signal source.

7.1.2.3 Rhythm

All three examples of rhythm were perceived by the majority of subjects as having a high arousal effect. This is one of the elements where it appears as though subjects struggled to identify valence accurately. As discussed in Chapter 6, the results of the rhythm examples were somewhat counter-intuitive as it would be expected that if the simplest rhythm was the happiest then the medium one should be second happiest and the complex rhythm should be the least happy. This, however, has not proved to be the case in the current research. The most complex rhythm rates as the happiest, the simple one second and the medium-complexity example third. The arousal scores confirm this as, once again, the simplest rhythm should rate as lowest arousal and the most complex as highest arousal, and yet this order is reversed. Another oddity here is that the medium-complexity rhythm was mostly perceived as aggressive whereas both the simple and complex ones were perceived mostly as happy.

Nonetheless, according to the data captured, the rhythms were all successful in evoking an effect which was the same for most respondents, especially if valence judgements are excluded. It also appears that rhythm, whether simple or complex, evokes a high arousal in listeners.

7.1.2.4 Peak Sound Level

Each of the four examples of peak sound level that was tested in the listening experiment was rated as aggressive by a majority of respondents. A few other respondents perceived it as happy. When these two sets of responses are combined for the peak sound level tests and compared with responses for calm/melancholic, it is clear that no matter how fast or slowly it occurs a volume change of 30dB evokes a high arousal-state in listeners.

7.1.2.5 Repetition

The testing of this element was very interesting as the question of whether a listener could discern multiple repetitions of a random segment of a noise signal was intriguing to begin with. The data collected here did not disappoint, and clearly shows that repetitions at intervals of 500ms or more were perceived as low arousal, and 250ms or less were perceived as high arousal. From this data it is clear that the brain can perceive these repetitions and respond to them. One question that arises here is: what is the dividing line between high and low arousal regarding repetition? According to the data, it lies somewhere between 500ms and 250ms. This could be an opportunity for future research.

So with the above in mind the three elements we can describe as successful in consistently evoking the musical effect within the limited scope of this research were repetition, peak sound level and rhythm. The remaining two elements tested, spectral shape and dynamics, could be classed as moderately successful. The results produced were interesting and compelling at certain points – but inconclusive in that some elements within the results obtained in response to these questions make further research necessary in order to categorically state their efficacy.

7.1.3 Which properties of digital signal processing methods are best suited for use in the creation of noise signals which evoke the musical effect?

Having discussed the musical elements which were most effective in evoking the musical effect, our attention now turns to the Digital Signal Processing techniques which were used in generating these musical elements. It is logical that the most effective DSP techniques must be linked to the elements which were most evocative. As such a discussion of the techniques implemented to generate the three most successful musical elements follows.

7.1.3.1 Delay

The musical element of repetition was generated through the implementation of a delay line. As per Chapter 4 a delay line is used to manipulate the time-axis of a signal by making copies of the initial signal, and then holding the copies and releasing them at the desired time-intervals (Pohlman 2000:631,632; Case 2007:207). In terms of the element created for the listening experiment the delays times used would all be classified as long delays as they ranged between 125ms and 2 seconds in length and a long delay is classified as any delay longer than 60ms – see Chapter 4.

7.1.3.2 Volume Control

The peak sound level examples in the listening test were generated by the use of one of the most basic operations of DSP mixing – a volume controller. Volume was taken from a nominal beginning level of 65dB SPL and increased via the volume controller on the mixing console of *Reason* to 95dB SPL.

7.1.3.3 Equalizer (digital filter)

Rhythms in this listening experiment were created using one of the basic functions of a digital filter: a gain controller. Specifically, the gain controller which controls the amount of boost or cut on a two-band equalizer was implemented and boosted at time-intervals consistent with the note-values in the three rhythmic patterns to be created. The 'Q' or bandwidth was set as narrow as possible so as to generate a small, mid-frequency boost at the arrival of each note-value in the rhythm.

Although the DSP techniques employed in generating the dynamic and spectral shape tests were not quite as successful as the three above, I feel that they achieved results substantial enough to warrant a brief overview of the method used, as another researcher could potentially implement these techniques in order to attempt to achieve a greater degree of success than has been achieved here.

7.1.3.4 Volume controller with automation

The dynamic examples in the listening experiment were created through the use of a volume controller automated to oscillate according to the parameter (abrupt or continuous) assigned to it. Another option for the control of the volume here would be to utilize an external function such as the fluctuation of a Low Frequency Oscillator from a synthesizer to control the volume of the audio.

7.1.3.5 Equalizer (filter)

The spectral shape questions were generated through the use of an equalizer – an implementation of filtering – to boost or cut certain frequencies. For the 'bright' and 'dark' spectral shapes a high shelf and low shelf, respectively, were used. For the 'mid-heavy' and 'mid-lite' spectral shape examples a boost and cut with a set bandwidth were employed.

7.1.4 What are the possible applications of these techniques in both the music world and other spheres?

Having ascertained from the data gathered from the listening experiments in the current research that noise can be used to evoke the musical effect with some degree of success, and having used the data to investigate which DSP techniques were most evocative of this musical effect, what are the possible uses for DSP-manipulated noise in the world of music and other spheres? The ideas to follow assume that the data captured in the listening tests is true for a broader population. However, as stated in previous chapters, this may not be the case due to the small sample of respondents and lack of diversity in terms of cultural background and other factors.

7.1.4.1 Musical applications:

Many composers of music constantly search for new avenues of self-expression and new sounds with which to work. Noise manipulated using DSP techniques such as the ones above could become one such new medium within which to compose. In the current research at least three successful manipulation tools have been uncovered which could be used by composers to evoke emotion in their listeners using sounds which are unfamiliar as musical material. Further investigation into this subject matter may well reveal many more potential tools such as the ones uncovered here. These could be used within the context of an ensemble playing traditional instruments to augment the sound-palette or, for something more progressive, an ensemble of 'noise-musicians' could be formed, performing entire compositions of white noise music.

Another interesting musical application for the knowledge gain suggested by this research lies in the field of music education. Due to the fact that noise – a non periodic, random signal – was used as a signal source to be manipulated via DSP techniques to try to evoke the musical effect (which, as defined in Chapter 3 is the emotional response evoked by music in the listener), many of the more traditional musical media such as notes, chords and metre were removed as the means to evoke emotion. These more common musical tools are the ones which traditional music education focuses on (sometimes almost exclusively). However, the findings of the current research suggest that there may be other parts of the sonic whole which we call music that have power to evoke the musical effect, and as such, should be taught to students of music as equal in importance to such elements as notes and rhythm. Education in musical elements such as spectral shape, peak sound level and repetition and the effect that each one evokes would surely benefit music students by availing them of new avenues for creativity.

7.1.4.2 Other applications:

When one considers the implications of noise being a potential evoking agent of the musical effect, the possibilities for the uses of white noise become fairly vast. I do not propose to mention every single possibility for its use here, but rather to mention the areas in which I feel it could be utilized.

For some time now, white noise has been used as a calming agent in devices such as sleep-aids, in that it masks outside noise so as to provide a quiet sleep-environment. An immediate possibility here is that this unaffected white noise could be treated in such a way as to induce a calming effect (possibly using the small dynamic range technique from the listening experiment) at the same time as being used as a masking agent thus improving its efficacy as a sleep-aid. In any situation where one wishes to calm people in a public space, this technique could be employed by public address system broadcast. Possible examples are shopping centres, post offices and waiting rooms.

Sound-design in movies could benefit from the use of noise in place of more traditional musical elements as the noise could be positioned at a low volume level and manipulated to produce the desired effect.

The field of music therapy could seek to exploit the evocative power of noise as an alternative to using only music as a therapeutic tool.

In any environment where privacy is needed, such as the offices of a counselling psychologist, manipulated noise could be broadcast in the waiting room as a calming measure and simultaneously be used to mask the sound of conversation from the counselling venue.

In recent times there has also been a move toward using extremely loud, highly directional sound by the military as a weapon in combat situations and where crowd control is needed. Sound recordings have also been used to disorientate or confuse enemies in combat. Further research could be undertaken into the possible use of DSP manipulated white noise in the same situations.

7.2 CONCLUDING REMARKS

This treatise set out to investigate whether or not noise could be used to evoke the musical effect when manipulated via Digital Signal Processing techniques. I acknowledge that no claim of universal truth can be made from an experiment which only received 31 responses and was limited to an unspecified portion of the population with limited age and cultural diversity. In order to validate these results the experiments could be repeated with a much larger sample group and the other factors above taken into account. However, evidence collected from both the literature study and listening experiment suggests that the answer to this question is 'Yes'. Various musical elements were represented via digital techniques and in the majority of cases respondents reported perceptions which conclusively indicated that some effect was felt. The techniques implemented and musical elements represented have been discussed, and possible applications for this knowledge have been explored. Further research could also be conducted into the implementation of even more DSP techniques, and also into garnering a more specific idea of the emotion perceived by respondents in response to the experiment material. Also, an investigation into the role of expectation and prediction in determining response and emotion – discussed on page 28 – when a manipulated noise signal is presented may be interesting and beneficial.

Some other interesting findings in the current research which pose questions for further investigation are:

1. Volume change – especially large volume change – was one of the most effective techniques employed in the listening tests. Yet in recent years this has become a vastly under-utilized element of popular music, due to the increased compression used by audio engineers to satisfy the desire of artists to produce the 'loudest' song possible (so that when people listen to it on a portable music player it stands out). The result of this compression is that the dynamic range of popular music is getting smaller and smaller. From the results obtained in this experiment one wonders what the affective cost to the music has been of this decrease in dynamic range.

2. By far the fewest of the responses to the test material were to high arousal and positive valence or 'happy'. Generally when respondents felt high arousal they selected 'aggressive'. Many respondents remarked after they completed their tests that they found the noise annoying, irritating and unpleasant. The question which presents itself here is: Is white noise generally perceived as an unpleasant sound? If so, is this due to its randomness? Or is it some negative association which people have to white noise?

3. Repetition is a key element in music. It appeals to the listener's desire to predict what is to happen next. The idea to test whether or not respondents would perceive any effect in response to continuous repeated noise was an intriguing one as it would indicate that the human brain can perceive patterns in randomness if they are repeated a sufficient number of times. The data captured from these questions is very clear that respondents could recognise that there were repetitions in the audio. Even more interesting in the data is that there seems to be a defining time value beyond which the perceived arousal value changes. Repeats of longer than 500ms

were all classified as calm/melancholic, and shorter than 250ms as aggressive. Two questions arise here: what is the exact value between 250ms and 500ms at which the perception change occurs? And what is the basis for these perceptions?

4. In the tests of spectral shape the questions for 'dark' spectral shape was responded to overwhelmingly as aggressive. All of the rest were either undefined, or calm/melancholic. The question that presents itself here is: Is there perhaps some primal association with a dark spectral shape which makes people feel highly aroused and negative (one thinks here of sounds in nature which heed as warning-signs to seek safety such as thunder or the roar of a large predator)?

In conclusion, the data collected in this study suggests that noise, manipulated via DSP techniques in real-time, can produce the musical effect. The current research has illuminated many points of interest and asked many questions in the area of the human perception of emotion in music, and it is evident that there is a need for further research into various facets of this topic. It is hoped that this treatise will encourage further research in this promising field.

SOURCES

8.1 BOOKS:

Adler, E. (2008). *How It's Done: An Invitation to Social Research*. Thomson Wadsworth. Belmont CA.

Alasuutari, P. Bickman, L. and Brannen,J. (eds) (2008). *The SAGE Handbook of Social Research Methods.* SAGE Publications. Los Angeles.

Babbie, E and Mouton, J. (2005). *The Practice of Social Research*. Cape Town: Oxford University Press

Babbie, E. and Mouton, J. (1998). *The Practice of Social Research*. Oxford, South Africa.

Ballou, G. (ed). (2002). *Handbook for Sound Engineers.* Oxford: Focal Press

Barthes, R. (1973). *Mythologies.* trans. Annette Lavers. New York: Hill and Wang.

Case, A. (2007). *Sound FX: Unlocking the Creative Potential of Recording Studio Effects*. Amsterdam. Focal Press.

Chase, W. (2006). *How Music REALLY Works!* second edition. Vancouver, Canada: Roedy Black Publishing.

Cook, P. (2002). *Real Sound Synthesis for Interactive Applications*. Natick, Mass. AK Peters.

Copland, A. (1963). *Copland on Music.* Pyramid Books. New York.

Creswell, J. (2009). *Research design.* California: Sage Publications

de Saussure, F. (1983). *Course in General Linguistics.* trans. Roy Harris. La Salle, Ill.: Open Court.

Deutsch, D. (ed). (1982). *The Psychology of Music*. Academic Press. New York.

Dodge, Y. (2003). *The Oxford Dictionary of Statistical Terms*. OUP.

Frith, S and Goodwin, A (eds) (1990): *On Record:Rock, Pop and the Written Word*. London: Routledge.

Hogg, R. V. and Tanis, E. A. (2005). *Probability and statistical inference*. (7th ed.). New Jersey, NJ: Prentice Hall.

Ifeachor, E and Jervis, B. (2002). *Digital Signal Processing: A practical Approach.* Pearson Education Ltd. 2nd ed. London, England.

Israel, M., and Hay, I. (2006). *Research ethics for social scientists.* London: Sage Publications

Juslin and Sloboda. (2001). *Music and Emotion.* Oxford. Oxford Press.

Katz, B. (2002). *Mastering Audio.* Oxford: Focal Press

Kefauver, A. (2001). *The Audio Recording Handbook.* Middletown, Wis. A-R Editions.

Kish, Leslie. (1965). *Survey Sampling.* New York: John Wiley and Sons, Inc.

Kivy, P. (1980). *The Corded Shell: Reflections on Musical Expression.* Princeton, NJ:Princeton Essays on the Arts.

Langer, S. (1942). *Philosophy in a new key*. Cambridge, M.A.: Harvard University Press.

Lerdahl, F and Jackendoff, R. (1983). *A Generative Theory of Tonal Music*. MIT Press.

Levitin, D. (2006). *This is Your Brain on Music.* Atlantic Books. London.

Macone, R. (1990). *The Concept of Music.* Oxford. Clarendon Press.

Mann, Prem S. (1995) *Introductory Statistics.* 2nd Edition, Wiley.

Mazzola, G. (2002). *The Topos of Music.* Birkhauser. Basel, Switzerland.

Meyer, L. (1956). *Emotion and Meaning in Music*. Phoenix Books. Chicago.

Mitchell, M. (2010). *Research Design Explained.* Wadsworth. 7th Ed. Belmont, California.

Mouton, J. (2001). *How to succeed in your Master's and Doctoral studies*. Pretoria: Van Schaik

Nardi, P. (2006). *Doing Survey Research: A guide to quantitative methods*. Pearson Education. Boston.

Nattiez, J. (1990). *Music and Discourse.* Princeton University Press. New Jersey.

Pohlman, K. (2000). *Principles of Digital Audio.* New York: McGraw Hill

Roads, C. (1996). *The Computer Music Tutorial.* Cambridge: The MIT Press

Roads et al. (eds). (1997). *Musical Signal Processing.* Lisse: Sweets

Roederer, J. (1979). *Introduction to the Physics and Psychophysics of Music*. Springer Verlag. New York.

Russ, M. (1996). *Sound Synthesis and Sampling.* Oxford: Focal Press

Russ, M. (2004). *Sound Synthesis and Sampling.* 2nd Ed. Amsterdam. Elsevier.

Russolo, L. (1913). *The Art of Noise: A futurist manifesto*

Sheperd and Wicke. (1997). *Music and Cultural Theory.* Polity Press. Cambridge.

Smith, F. (1979). *The Experiencing of Musical Sounds.* Gordon and Breach Science Publishers. New York.

Thayer, R. E. (1989). *The Biopsychology of Mood and Arousal.* New York: Oxford University Press.

Various. (2003). *Collins English Dictionary*. HarperCollins Publishers. London, England.

Various. (1988). *Webster's English Dictionary*. Melbourne: Budget books

Walsh, M. (2001). *Research made real: a guide for students.* Cheltenham: Nelson Thornes

Woods, A. (1975). *The Physics of Music*. 7th Ed. Chapman and Hall. London.

Xenakis, I. (1971). *Formalized Music.* Indiana University Press. London.

8.2 ARTICLES:

Bigand, E and Veillard, S. (2005). *Multidimensional scaling of emotional responses to music: The effect of musical expertise and of the duration of the excerpts*. Cognition and Emotion. Vol. 19. Psychology Press Ltd.

Bresin, R and Friberg, A. (1999). *Synthesis and Decoding of Expressive Music Performance*.

Stockholm, Sweden. Royal Institute of Technology. IEEE.

Gabrielsson, A and Juslin, P. (1996). *Emotional Expression in Music Performance: Between the Performer's Intention and the Listener's Experience*. Psychology of Music, Vol. 24, pp. 68-91.

Gabrielsson, A. and Lindström, W. (2003). *Strong experiences related to music: A descriptive system.* Musicae Scientiae, Vol. 7, pp. 157-217.

Gabrielsson, A., and Lindström, E. (2001). *The influence of musical structure on emotional expression.* In P. N. Juslin and J. A. Sloboda (Eds.), Music and emotion: Theory and research pp. 223-248. New York: Oxford University Press.

Goldman, Richard Franko. (1961). *Varèse: Ionisation; Density 21.5; Intégrales; Octandre; Hyperprism; Poème Electronique.* Instrumentalists, cond. Robert Craft. Columbia (in Reviews of Records). Musical Quarterly 47, no. 1. (January) pp.133–34.

Toussaint, G. Gomez, F. and Thul, E (2007*). An experimental comparison of formal measures of rhythmic syncopation.* In Proceedings of the International Computer Music Conference, pp. 101– 104, Aug 2007.

Grewe, O., Nagel, F., Kopiez, R., and Altenmüller, E. (2005). *How Does Music Arouse "Chills"? Investigating Strong Emotions, Combining Psychological, Physiological, and Psychoacoustical Methods.* Annals of the New York Academy of Science, 1060, pp. 446-449.

Grewe, O., Nagel, F., Kopiez, R., and Altenmüller, E. (2007). *Listening to Music as a Re-Creative Process: Physiological, Psychological, and Psychoacoustical Correlates of Chills and Strong Emotions*. Music Perception, 24, pp. 297-314.

Guhn, M., Hamm, A., and Zentner, M. (2007). *Physiological and Musico-Acoustic Correlates of the Chill Response.* Music Perception, 24, pp. 473-483.

Han, B. (2009). *Music emotion classification and context-based music recommendation.* Multimed Tools Appl (2010) 47:433–460. Published online: 5 August 2009. Springer Science + Business Media, LLC 2009

Juslin, P and Laukka, P. (2003). *Communication of Emotions in Vocal Expression and Music Performance: Different Channels, Same Code?* Psychological Bulletin Vol. 129, No. 5

American Psychological Association, Inc.

Juslin, P. (2000). *Cue Utilization in Communication of Emotion in Music Performance: Relating Performance to Perception*. Journal of Experimental Psychology: Human Perception and Performance. Vol. 26, No. 6, pp. 1797-1813

Kim, Y. et al. (2010). *Music Emotion Recognition: A State of the Art Review*. 11th International Society for Music Information Retrieval Conference (ISMIR 2010)

MacDorman, K. F et al. (2007). *Automatic Emotion Prediction of Song Excerpts: Index Construction, Algorithm Design, and Empirical Comparison*. Journal of New Music Research 2007, Vol. 36, No. 4, pp. 283–301.

McLean, B. (1981). *Symbolic Extension and its Corruption of Music.* Perspectives of New Music, Vol. 20, nos. 1-2

Min Xu et al. (2004). *HMM-based audio keyword generation.* In Kiyoharu Aizawa, Yuichi Nakamura, Shin'ichi Satoh. Advances in Multimedia Information Processing - PCM 2004: 5th Pacific Rim Conference on Multimedia. Springer.

Nagel, F., Kopiez, R., Grewe, O., and Altenmüller, E. (2008). *Psychoacoustical correlates of musically induced chills.* Musicae Scientiae, Vol. 12, pp. 101-113.

Paleari, M. et al. (2007). *Using Emotions to Tag Media.* Jamboree 2007: One Day Workshop By and For KSpace Ph.D. Students, Berlin, Germany.

Palmer, C. (1997). *Music performance.* Annual Review of Psychology, Vol. 48, pp. 115-138.

Patterson, R. D., Handel, S., Yost, W. A., and Datta, A. J. (1996). *The relative strength of the tone and noise components in iterated rippled noise.* Journal of the Acoustical Society of America. Vol. 100, pp. 3286–3294.

Sahidullah, Md. and Saha, G. (2012). *Design, analysis and experimental evaluation of block based transformation in MFCC computation for speaker recognition*. Speech Communication Vol. 54. pp. 543–565.

Seashore, H. G. (1937). *An objective analysis of artistic singing*. In C. E. Seashore (Ed.), University of Iowa studies in the psychology of music: Vol. 4 Objective analysis of musical performance. pp. 12-157. Iowa City: University of Iowa.

Sloboda, J. A. (1991). *Music Structure and Emotional Response: Some Empirical Findings*. Psychology of Music, Vol. 19, pp. 110-120.

Toussaint, G. (2002). *A mathematical analysis of African, Brazilian, and Cuban clave rhythms*. In BRIDGES: Mathematical Connections in Art, Music and Science, pp. 157–168, Jul 2002.

Trohidis, K et al. (2008). *Multi-label Classification of Music into Emotions.* ISMIR 2008 – Session 3a – Content-Based Retrieval, Categorization and Similarity

van de Laar, B. (2006). *Emotion detection in music, a survey.* Proceedings of the Twente Student Conference, IT.

Wasserman, K. Eng, K. And Verschure, P. (2003). *Live Soundscape Composition Based on Synthetic Emotions*. IEEE Multimedia, Vol. 10, pp. 82–90.

Wu, T-L. et al. (2006). *Automatic emotion classification of musical segments.* 9th international conference on music perception and cognition. Alma Mater Studiorum University of Bologna, August 22-26, 2006.

Yasuda, S. and Nakamura, T. (2008). *A psychological study of strong experiences due to listening to music based on a subjective measurement of physical reactions*. The Japanese Journal of Cognitive Psychology. Vol. 6. pp. 11-19.

Yost, W. (1997). *Pitch strength of iterated rippled noise when the pitch is ambiguous.* Journal of the Acoustical Society of America. Vol. 101(3). pp. 1644-8.

8.3 WEBSITES REFERENCED:

Bierce, A. (1911). *The Devil's Dictionary.* http://www.hermetics.org/pdf/Bierce,_Ambrose_-_The_Devil-s_Dictionary.pdf (accessed on 4 April 2013)

Cook, J. (2008). *Error in the normal approximation to the t distribution*. (accessed on 21 November 2012)

http://www.johndcook.com/normal_approx_to_t.html

Chang, JH. (1999). *Composing noise*. (accessed on 15 June 2007) http://www.interdisciplina.org/microsound-repository/index.php &direction=0&order=&directory=miscellaneous/Reading

Copland, A. (1939). *How we listen.* Accessed at http://www.montgomeryschoolsmd.org/uploadedFiles/schools/wheatonhs/departments/aphonors/Copland_HowWeListen.pdf

Dennis, R. (2001). *Advanced Recording Primer*. (accessed on 20 November 2012). http://www.recordinginstitute.com/da154/ARP/index.html

Freed, G. (1998).(moderator). *The state of classical music: A panel discussion.* Transcription for National Arts Journalism Program at Columbia University. (accessed on 16 June2007) http://www.najp.org/publications/occasionalreports.html

Gunes et al. (2011). *Emotion Representation, Analysis and Synthesis in Continuous Space: A Survey.* http://ibug.doc.ic.ac.uk/media/uploads/documents/gunesetal_emospace2011_4feb2011.pdf (accessed on 15 August 2012)

Hegarty, P. (2001) *Full With Noise: Theory and Japanese Noise Music* (accessed on 11 June 2007) http://www.ctheory.net/articles.aspx?id=314

Hu, J-X and Downie, J. S. (2010). *When Lyrics Outperform Audio for Music Mood Classification: A Feature Analysis*. 11th International Society for Music Information Retrieval Conference (ISMIR 2010). http://ismir2010.ismir.net/proceedings/ismir2010-106.pdf. (accessed on 18 June 2012)

Jacobs, D. (2005). *Correlation and Convolution.* http://www.cs.umd.edu/~djacobs/CMSC426/Convolution.pdf. (accessed on 15 Novemeber 2012)

Knapp, R and Cook, P. (2005). *The Integral Music Controller: Introducing a Direct Emotional Interface to Gestural Control of Sound Synthesis*. http://www.music.mcgill.ca/~ich/research/misc/papers/cr1316.pdf (accessed on 25 June 2012)

Martin, G. (2011). *Introduction to Sound Recording*. E-book. Version number 2011 09 11. http://www.tonmeister.ca/main/textbook/index.html. (accessed on 15 Novemebr 2012)

OECD. 2003. Business Tendency Survey Handbook. STATISTICS DIRECTORATE. from http://www.oecd.org/std/leadingindicatorsandtendencysurveys/33659904.pdf

Panda, R. and Paiva, R. (2011). *Automatic Creation of Mood Playlists in the Thayer Plane: A Methodology and a Comparative Study.* SMC Conference 2011. http://smcnetwork.org/smc_papers-2011156 (accessed on 18 June 2012)

Pohle, T. et al. (2005). *Evaluation of Frequently Used Audio Features for Classification of Music into Perceptual Categories.* Austrian Research Institute for Artificial Intelligence, Vienna. www.cp.jku.at/research/papers/Pohle_CBMI_2005.pdf (accessed 18 June 2012)

Ruckman, C. (2007). *Active noise Control FAQ.* http://www.chrisruckman.com/Active%20Noise%20Control%20FAQ%202007-02-04.txt (accessed on 4 June 2007)

Sangild, T. (2002). *The Aesthetics of Noise*. www.ubu.com/papers/noise.html (accessed on 15 June 2007)

Tagg, P. (1999). *Introductory notes to the Semiotics of Music*. Version 3: Liverpool/Brisbane, July 1999. http://www.tagg.org/xpdfs/semiotug.pdf (accessed on 6 September 2012)

Thul, E. (2008). *Measuring the Complexity of Musical Rhythm.* McGill University. Montreal, Quebec. http://www.cgrl.cs.mcgill.ca/~godfried/teaching/mir-reading-assignments/Eric-Thul-Thesis.pdf (accessed on 22 October 2012)

Tolos, M. Tato, R. and Kemp, T. (2005). *Mood-based navigation through large collections of musical data.* SONY International. Stuttgart, Germany. http://jan.newmarch.name/conferences/ccnc05/DATA/3-A01-02.PDF (accessed on 15 October 2012)

Trohidis et al. (2011). *Multi-label classification of music by emotion.* EURASIP Journal on Audio, Speech, and Music Processing 2011, 2011:4. http://asmp.eurasipjournals.com/content/2011/1/4

Weil, A. (2008). *Can Noise Make You Sick?* http://www.drweil.com/drw/u/QAA400439/Can-Noise-Make-You-Sick.html (accessed on 4 April 2013).

Yang, Y-H; Lin, Y-C; Su, Y-F and Homer H. (2008). *A Regression Approach to Music Emotion Recognition*. mpac.ee.ntu.edu.tw/~yihsuan/pub/TASLP08.pdf. (accessed on 18 June 2012)

Yang, Y-H; Yi-Hsuan Yang; Yu-Ching Lin; Heng-Tze Cheng; I-Bin Liao; Yeh-Chin Ho; and Homer H. Chen. (2008). *Toward Multi-Modal Music Emotion Classification*. National Taiwan

University. mpac.ee.ntu.edu.tw/~yihsuan/pub/PCM08.pdf. (accessed 18 June 2012)

Yang, Y-H; Yi-Hsuan Yang, Ya-Fan Su, Yu-Ching Lin, and Homer H. Chen (2007). *Music Emotion Recognition: The Role of Individuality.* Graduate Institute of Communication Engineering, National Taiwan University. mac.iis.sinica.edu.tw/pub/HCM07.pdf. (accessed on 18 June 2012)

"Digital" (accessed on 2007/06/22)

URL: http://encarta.msn.com/dictionary_1861604662/digital.htmlb

"Analog" (accessed on 2007/06/22)

URL: http://encarta.msn.com/dictionary_/analog.html

APPENDICES

9.1 APPENDIX A - PERMISSION TO APPROACH STAFF AT HARVEST CHRISTIAN SCHOOL AS TEST SUBJECTS

Hunt Street

Richmond Hill

Port Elizabeth

6006

18 October 2012

Dear Mr. Brettenny

Re: Master's Degree Experiments on Friday 9 November 2012

I am currently approaching the experimental phase of my Master's Degree in Music Technology and write to ask for your permission to include the Harvest Christian School staff in my experiments.

The experiments are perceptual in nature. Subjects will be played various sounds and asked to respond to what they hear by answering a questionnaire. They will not be exposed to anything unsafe, and will at all times retain the option to opt out of the test if they feel uncomfortable in any way.

Each experiment should take no more than 20 minutes. I plan to use my office at Harvest Christian School as a testing venue, and would propose to conduct the experiments between 07h30 and 17h00 from Wednesday 7 to Thursday 8 November 2012.

Your permission and support in this matter would be greatly appreciated as I feel it would help me to garner maximum participation in this crucial part of my project.

If you have any queries or concerns please feel free to contact me on 082 306 6967, or andrew.warneke@gmail.com.

Yours sincerely

Andrew Warneke

9.2 APPENDIX B - NOTICE TO STAFF AT HARVEST CHRISTIAN SCHOOL TO CREATE AWARENESS OF EXPERIMENTS

ATTENTION STAFF:

I will be conducting my experiments for my Master's Degree in Music Technology from 7 to 8 November from 07h30 'til 17h00 in my office in the music house.

It would help me greatly if you would give twenty minutes of your time to be part of this process as I require a reasonable number of respondents in order to have a good data set to analyse.

The experiment is a perceptually based listening test, and will not put you or your identity at risk in any way, and you will be free to opt out at any time if you so choose.

If you are able to make it please let me know. I can be contacted at 082 306 6967 or andrew.warneke@gmail.com

Many thanks

Andrew Warneke

9.3 APPENDIX C- PRETEST INFORMATION TO BE COMMUNICATED TO SUBJECTS

Information to be presented to listening test subjects prior to commencement of listening tests

Researcher: Andrew Warneke

1. Welcome

2. General background to project: the project seeks to evaluate the effectiveness of selected DSP techniques in evoking the musical effect

3. Explanation of nature of the questionnaire to be completed

4. Explanation of nature of listening tests with reference to the idea that the tests are perceptual in nature and that there are thus no 'right' or 'wrong' answers, only individual opinions

5. Information regarding the subjects' right to opt out of the test at any time for any reason

6. Assurance of anonymity of subjects outside of the testing environment

9.4 APPENDIX D - PRETEST LETTER TO BE GIVEN TO SUBJECTS

6 Hunt Street

Richmond Hill

Port Elizabeth

6006

5 November 2012

Dear Respondent

Thank-you for your willingness to participate in the experiment phase of my research study as part of my Master's Degree in Music Technology course. Your taking-part will help to make this endeavour of mine a success, and I value greatly your input and the time you have agreed to give.

This project seeks to evaluate the effectiveness of selected Digital Signal Processing techniques in evoking the 'musical effect' – which is an emotional response in the listener.

Please mark the option **you feel** best describes each piece of audio you hear after the voice prompts, as the experiment material is perceptual in nature, and thus, there is no right or wrong answer to any of the questions.

Please feel free to ask questions regarding the experiments before or after the testing has taken place. Also, please ensure that you are comfortable in the testing environment and that you will not be distracted by your cellphone during the testing process as the test material is continuous and does not provide breaks or pauses in which such distractions could be rectified.

Note that it is your right to leave the experiment at any point if you feel uncomfortable or experience anything that does not meet with your approval. Precautions have however been taken to attempt to exclude this possibility.

Your identity will remain private and will not be used for any purpose outside of this research. You will also not write your name on your questionnaire so as to ensure that all feedback is anonymous.

If you would like to be informed of the outcome of the research or have access to the completed document please email me at andrew.warneke@gmail.com and I will make the information available to you when possible.

Thanks again for your interest, willingness and time.

Yours sincerely

Andrew Warneke (Researcher)

9.5 APPENDIX E - REC-H APPLICATION FORM (INFORMED CONSENT FORM COMPLETED BY ALL SUBJECTS)

Nelson Mandela Metropolitan University

for tomorrow

APPLICATION FOR APPROVAL
NMMU RESEARCH ETHICS COMMITTEE (HUMAN)

SECTION A: (To be filled in by a representative from the Faculty RTI Committee)

Application reference code:	H HUMAN	…………. YEAR	…………. FACULTY	…………. DEPARTMENT	…………. NUMBER

Resolution of FRTI Committee:	E R
Resolution date:	
Faculty RTI representative signature:	

BEFORE YOU FILL IN THIS FORM PLEASE READ THE FOLLOWING DOCUMENTS:

- "Research Ethics (Human) Application Process" (http://www.nmmu.ac.za/default.asp?id=4619&bhcp=1)
- "Code of Conduct for Researchers at NMMU" (Students: http://portal.nmmu.ac.za/default.asp?id=71&sp=0&bhcp=1 or Staff: http://my.nmmu.ac.za/default.asp?id=308&bhcp=1).

WHO NEEDS TO FILL THIS FORM IN?

Any project in which humans are the subjects of research (hereafter called a *study*) requires completion of this form and submission for approval first to their Faculty RTI Committee (FRTI). The FRTI will refer projects to the Research Ethics Committee (Human) (REC-H) where deemed necessary.

WHEN SHOULD THIS FORM BE HANDED IN?

The research proposal should first have been approved by the FRTI before Ethics approval may be given. It should also have first been reviewed by the FRTI for **Ethics** clearance before it is referred to the REC-H.

HOW TO FILL THIS FORM IN:

1) Complete Sections 1 to 8 in typescript (Tab between fields, select from pull-downs, information may be pasted from existing Word® documents), and save (filename must contain your name). Handwritten forms will not be accepted.
2) Use the "Save as" option to save the application form with a filename containing your name (e.g. "**J Smith** REC-H Application Form.doc").
3) Complete Sections 1 to 8 in typescript (Tab between fields, select from pull-downs, information may be pasted from existing Word® documents), and save (filename must contain your name). Handwritten forms will not be accepted.
4) Append the necessary information e.g. Research methodology, Informed consent form, Written information given to participant prior to participation, Oral information given to participant prior to participation (examples of these may be found on the Research Ethics webpage: (http://www.nmmu.ac.za/default.asp?id=4619&bhcp=1)
5) **Electonic copy:** Email all the files (including any appendices) to the Faculty RTI Committee representative in the relevant Faculty.

6) **Hard copy, signed:** Print the document, get each page initialled on the lower right hand corner and get Sections 9 and 10 signed by the relevant parties. **Hand the signed hardcopy and attachments in** to the Faculty RTI Committee representative in the relevant Faculty.

Please delete this instruction block before you save and print.

1. GENERAL PARTICULARS

TITLE OF STUDY

a) Concise descriptive title of study (must contain key words that best describe the study): Evaluation of the efficacy of digital real-time noise control techniques in evoking the musical effect

PRIMARY RESPONSIBLE PERSON (PRP)

b) Name of PRP (must be member of permanent staff. Usually the supervisor in the case of students): Dr Rudi Bouwer and Mr. Mark Brand

c) Contact number/s of PRP:

d) Affiliation of PRP: Faculty: Arts
Department (or equivalent): Music

PRINCIPLE INVESTIGATORS AND CO-WORKERS

e) Name and affiliation of principal investigator (PI) / researcher (may be same as PRP): Andrew Warneke
Gender: Male

f) Name(s) and affiliation(s) of all co workers (e.g. co-investigator/assistant researchers/supervisor/co-supervisor/promoter/co-promoter). If names are not yet known, state the affiliations of the groups they will be drawn from, e.g. Interns/M-students, etc. and the number of persons involved: none

STUDY DETAILS

g) Scope of study: h

i) Funding :
Additional information (e.g. source of funds or how combined funding is split)

j) Are there any restrictions or conditions attached to publication and/or presentation of the study results? No
If YES, elaborate (Any restrictions or conditions contained in contracts must be made available to the Committee):

k) Date of commencement of data collection: 15 June 2012
Anticipated date of completion of study: 1 September 2012

l) Objectives of the study (the major objective(s) / Grand Tour questions are to be stated briefly and clearly): Is it possible to use Digital Signal Processing techniques to manipulate noise and thereby evoke musical effects? What are the specific techniques that facilitate this end?

m) Rationale for this study: briefly (300 words or less) describe the background to this study i.e. why are you doing this particular piece of work. A few (no more than 5) key scientific references may be included: The beginnings of the selection of this topic of study emerged from a discussion I had with a business executive who consults to various industrial plants. He approached me regarding the problem of excessive noise in the factories he consults to, as these companies are forced to spend a large sum of money annually to purchase hearing protection for employees who work in close proximity to machinery, as the noise-levels are above the legal limit. He wondered whether there was not a way of using digital technology to reduce the noise-level, as acoustic control methods are rendered ineffectual due to the porous nature of the sound absorption materials (they become saturated with dust, and then only serve to reflect noise, instead of absorb it). Unfortunately, the technologies which could be used in this situation are still in their infancy, and research is still in progress as to how to expand Active Noise Control (Ruckman 2007) systems into a complex, visceral, sound-environment (Ruckman 2007). However, my discipline being music, this conversation did lead me on a path of creative thought regarding the use of noise in a creative manner – this study will purport to be the culmination of this creative thought process, together with an in-depth literature study, and an experimental study to test the effectiveness of the techniques uncovered.

METHODOLOGY

a) Briefly state the methodology (specifically the procedure in which human subjects will be participating) (the full protocol is to be included as *Appendix 1*): Human subjects will be involved in a listening test wherein they will audition various noise manipulation techniques and report – on a questionnaire – the effectiveness of the techniques in demonstrating the presence of musical effects

b) State the minimum and maximum number of participants involved (Minimum number should reflect the number of participants necessary to make the study viable) Minimum: 50 Maximum: 70

2. RISKS AND BENEFITS OF THIS STUDY

a) Is there any risk of harm, embarrassment or offence, however slight or temporary, to the participant, third parties or to the community at large? Yes. Complete anonymity is not guaranteed to subjects as they will all be present at the listening tests at the same time. The researcher does not see this as a major problem though, because measures will be taken to ensure that subjects cannot see the responses offered on the questionnaires of other participants, and as such, no harm or embarrassment should be incurred by any subject.

b) Has the person administering the project previous experience with the particular risk factors involved? No.
If YES, please specify:

c) Are any benefits expected to accrue to the participant (e.g. improved health, mental state, financial etc.)? **No.**

If YES, please specify the benefits:

d) Will you be using equipment of any sort? Yes
If YES, please specify: **A public address system, a computer, and a MIDI controller**

e) Will any article of property, personal or cultural be collected in the course of the project? No.
If YES, please specify:

3. TARGET PARTICIPANT GROUP

a) If particular characteristics of any kind are required in the target group (e.g. age, cultural derivation, background, physical characteristics, disease status etc.) please specify: None

b) Are participants drawn from NMMU students? No

c) If participants are drawn from specific groups of NMMU students, please specify:

d) Are participants drawn from a school population? No
If YES, please specify:

e) If participants are drawn from an institutional population (e.g. hospital, prison, mental institution), please specify:

f) If any records will be consulted for information, please specify the source of records:

g) Will each individual participant know his/her records are being consulted?
If YES, state how these records will be obtained:

h) Are all participants over 18 years of age? Yes
If NO, state justification for inclusion of minors in study:

4. CONSENT OF PARTICIPANTS

a) Is consent to be given in writing? Yes
If YES, include the consent form with this application [Appendix 2].
If NO, state reasons why written consent is not appropriate in this study.

b) Are any participant(s) subject to legal restrictions preventing them from giving effective informed consent? No
If YES, please justify:

c) Do any participant(s) operate in an institutional environment, which may cast doubt on the voluntary aspect of consent? **No**
If YES, state what special precautions will be taken to obtain a legally effective informed consent:

d) Will participants receive remuneration for their participation? No
If YES, justify and state on what basis the remuneration is calculated, and how the veracity of the information can be guaranteed.

e) Which gatekeeper will be approached for initial permission to gain access to the target group? (e.g. principal, nursing manager, chairperson of school governing body) None required

f) Do you require consent of an institutional authority for this study? (e.g. Department of Education, Department of Health) No
If YES, specify:

5. INFORMATION TO PARTICIPANTS

α) What information will be offered to the participant before he/she consents to participate? (Attach written information given as [*Appendix 3*] and any oral information given as [*Appendix 4*])

β) Who will provide this information to the participant? (Give name and role) Andrew Warneke, researcher

χ) Will the information provided be complete and accurate? **Yes**
If NO, describe the nature and extent of the deception involved and explain the rationale for the necessity of this deception:

6. PRIVACY, ANONYMITY AND CONFIDENTIALITY OF DATA

a) Will the participant be identified by name in your research? No
If YES, justify:

b) Are provisions made to protect participant's rights to privacy and anonymity and to preserve confidentiality with respect to data? Yes
If NO, justify. If YES, specify: Names will not be written on questionnaires

c) If mechanical methods of observation be are to be used (e.g. one-way mirrors, recordings, videos etc.), will participant's consent to such methods be obtained? Not used
If NO, justify:

d) Will data collected be stored in any way? Yes
If YES, please specify: (i) By whom? (ii) How many copies? (iii) For how long? (iv) For what reasons? (v) How will participant's anonymity be protected? The (i) researcher will store the questionnaires in (ii) hard-copy duplicate for (iii) 6 months (iv) in case reflection on the original data is required. (v) Anonymity is protected as names will not appear on the questionnaires.

e) Will stored data be made available for re-use? No
If YES, how will participant's consent be obtained for such re-usage?

f) Will any part of the project be conducted on private property (including shopping centres)? Yes
If YES, specify and state how consent of property owner is to be obtained: A letter will be written to the owners of

the property to ask permission to use the premises.

g) Are there any contractual secrecy or confidentiality constraints on this data? No
If YES, specify:

7. FEEDBACK

a) Will feedback be given to participants? **No, but they will be allowed to ask questions after the listening tests, and be given contact details of the researcher in order to request feedback at a later stage.**
If YES, specify whether feedback will be written, oral or by other means and describe how this is to be given (e.g. to each individual immediately after participation, to each participant after the entire project is completed, to all participants in a group setting, etc.):

b) If you are working in a school or other institutional setting, will you be providing teachers, school authorities or equivalent a copy of your results?
If YES, specify, if NO, motivate:

8. ETHICAL AND LEGAL ASPECTS

The Declaration of Helsinki (2000) or the Belmont Report will be included in the references:
If NO, motivate:
(A copy of the Belmont Report is available at the following link for reference purposes: http://www.nmmu.ac.za/documents /rcd/The%20Belmont%20Report.pdf)

a) I would like the REC-H to take note of the following additional information:

9. DECLARATION

If any changes are made to the above arrangements or procedures, I will bring these to the attention of the Research Ethics Committee (Human). I have read, understood and will comply with the *Guidelines for Ethical Conduct in Research and Education at the Nelson Mandela Metropolitan University* and have taken cognisance of the availability (on-line) of the Medical Research Council Guidelines on Ethics for Research (http://www.sahealthinfo.org/ethics/).
All participants are aware of any potential health hazards or risks associated with this study.
I aware of potential conflict(s) of interest which should be considered by the Committee.
If affirmative, specify:

20 November 2012

SIGNATURE: (Primary Responsible Person) Date

20 November 2012

SIGNATURE: (Principal Investigator/Researcher) Date

10. SCRUTINY BY FACULTY AND INTRA-FACULTY ACADEMIC UNIT

This study has been discussed, and is supported, at Faculty and Departmental (or equivalent) level. This is attested to by the signature below of a Faculty (e.g. RTI) and Departmental (e.g. HoD) representative, neither of whom may be a previous signator.

NAME and CAPACITY (e.g. HoD)	SIGNATURE	Date
NAME and CAPACITY (e.g. Chair:FacRTI)	SIGNATURE	Date

11. APPENDICES
In order to expedite the processing of this application, please ensure that all the required information, as specified below, is attached to your application. Examples of some of these documents can be found on the Research Ethics webpage (http://www.nmmu.ac.za/default.asp?id=4619&bhcp=1). You are not compelled to use the documents which have been provided as examples – they are made available as a convenience to those who do not already have them available.
APPENDIX 1: Research methodology
Attach the full protocol and methodology to this application, as "Appendix 1" and include the data collection instrument e.g. questionnaire if applicable.
APPENDIX 2: Informed consent form
If no written consent is required, motivate at 4a). The intention is that you make sure you have covered all the aspects of informed consent as applicable to your work.
APPENDIX 3: Written information given to participant prior to participation
Attach as "Appendix 3". The intention is that you make sure you have covered all the aspects of written information to be supplied to participants, as applicable to your work.
APPENDIX 4: Oral information given to participant prior to participation
If applicable, attach the required information to your application, as "Appendix 4".
APPENDIX 5, 6, 7: Institutional permissions
Attach any institutional permissions required to carry out the research e.g. Department of Education permission for research carried out in schools.

9.6 APPENDIX F - ORDER OF MUSICAL ELEMENTS & DSP TECHNIQUES IMPLEMENTED IN LISTENING TESTS

Order of testing musical elements and DSP techniques:

1. Dynamics – small range; continuous change
2. Spectral shape – dark
3. Rhythm – medium complexity
4. Peak Sound level – medium
5. Repetition – 2 sec
6. Rhythm – simple
7. Spectral shape – mid heavy
8. Peak Sound level – sudden
9. Dynamics – large range; discrete change
10. Repetition – 500ms
11. Repetition – 1 second
12. Peak Sound level – slow
13. Dynamics – large range; continuous change
14. Rhythm – high complexity
15. Spectral shape – bright
16. Repetition – 125ms
17. Peak Sound level – fast
18. Repetition – 750ms
19. Dynamics – small range; discrete change
20. Spectral shape - mid-lite
21. Repetition – 250ms

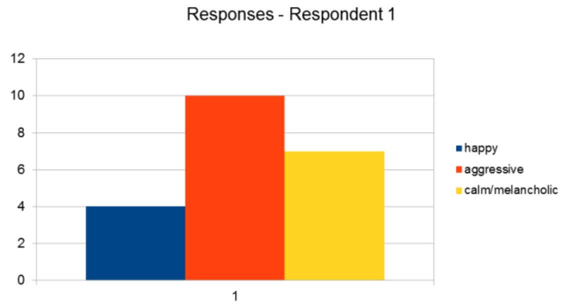
9.7 APPENDIX G – QUESTIONNAIRE USED IN LISTENING EXPERIMENTS

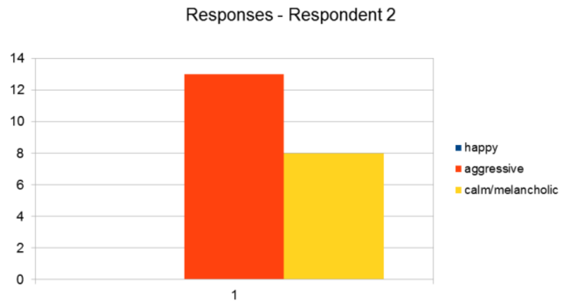
Listening Experiment Questionnaire

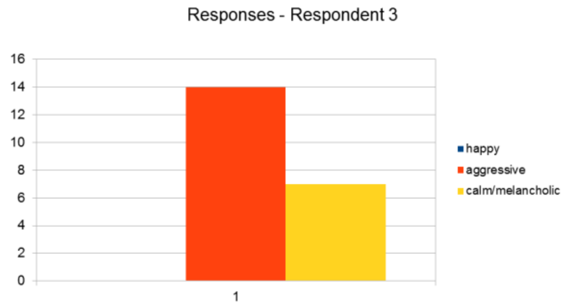
1. happy	aggressive	calm/melancholic
2. happy	aggressive	calm/melancholic
3. happy	aggressive	calm/melancholic
4. happy	aggressive	calm/melancholic
5. happy	aggressive	calm/melancholic
6. happy	aggressive	calm/melancholic
7. happy	aggressive	calm/melancholic
8. happy	aggressive	calm/melancholic
9. happy	aggressive	calm/melancholic
10. happy	aggressive	calm/melancholic
11. happy	aggressive	calm/melancholic
12. happy	aggressive	calm/melancholic
13. happy	aggressive	calm/melancholic
14. happy	aggressive	calm/melancholic
15. happy	aggressive	calm/melancholic
16. happy	aggressive	calm/melancholic
17. happy	aggressive	calm/melancholic
18. happy	aggressive	calm/melancholic
19. happy	aggressive	calm/melancholic
20. happy	aggressive	calm/melancholic
21. happy	aggressive	calm/melancholic

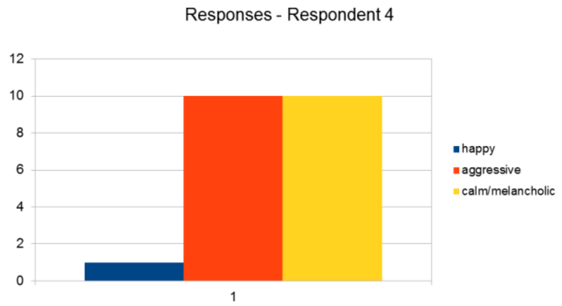
9.8 APPENDIX H - ADDITIONAL GRAPHS GENERATED FROM RAW DATA

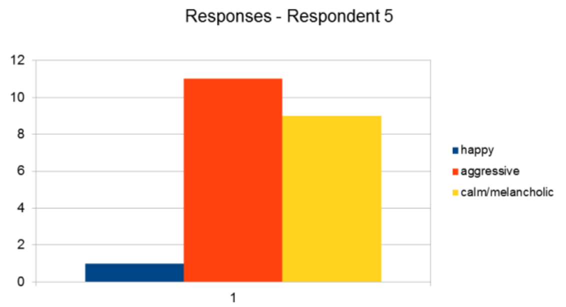
9.8.1 Responses per respondent for all questions

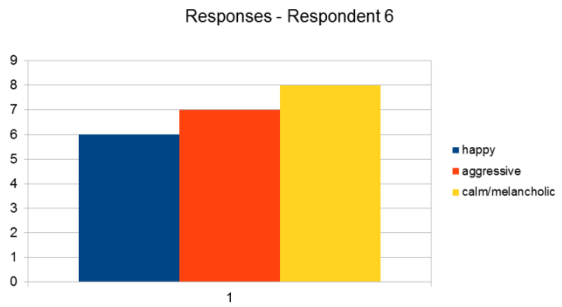


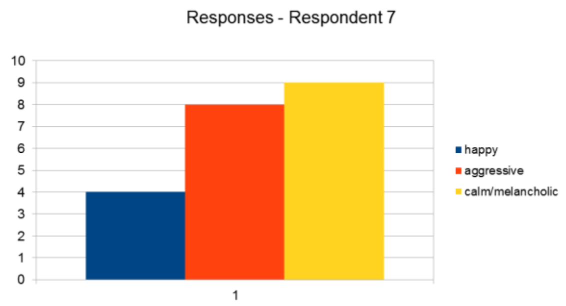


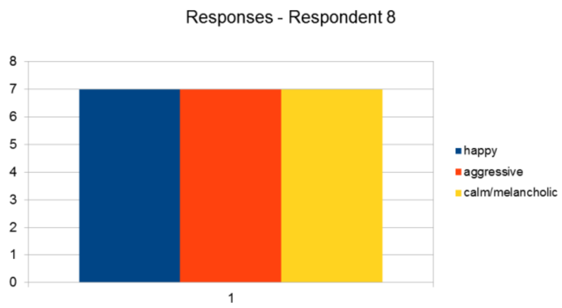


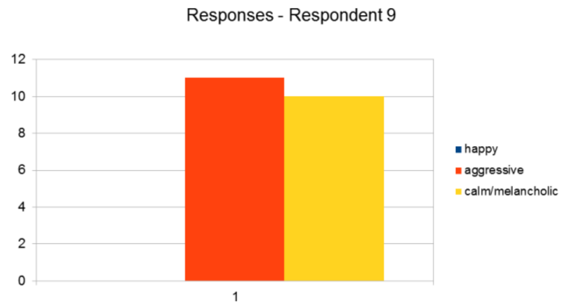


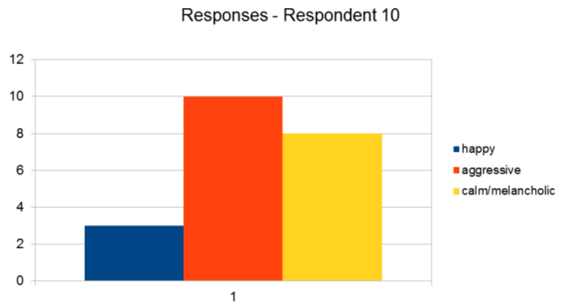


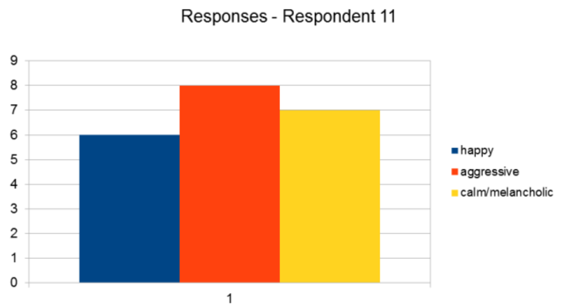


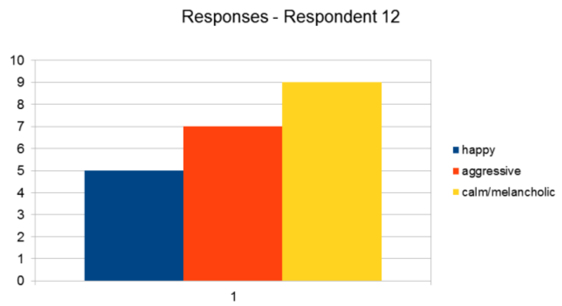


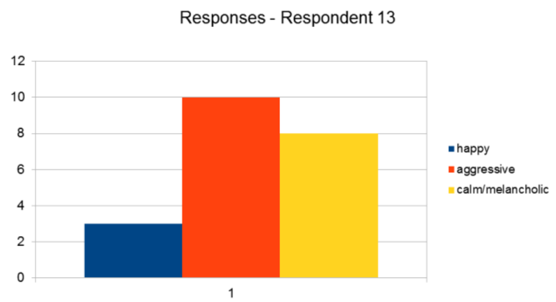


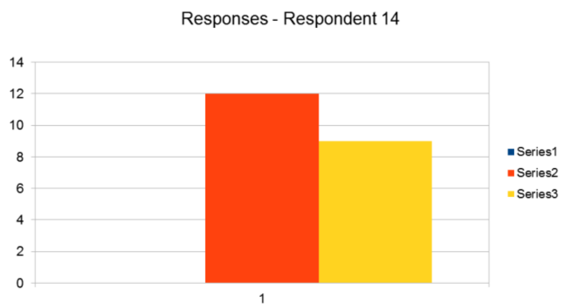

Responses - Respondent 9
12
10
8
6
4
2
0
1
happy
aggressive
calm/melancholic

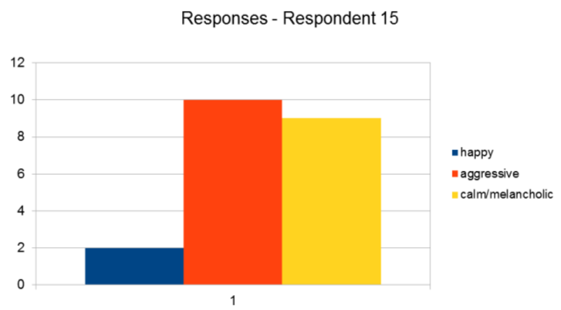

Responses - Respondent 10
12
10
8
6
4
2
0
1
happy
aggressive
calm/melancholic

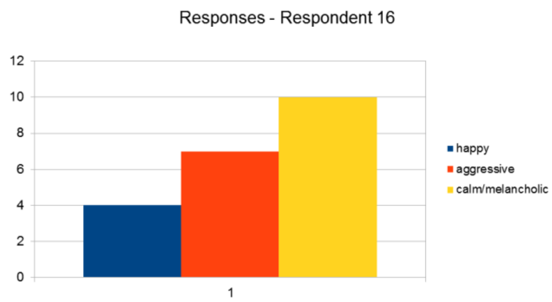

Responses - Respondent 11
9
8
7
6
5
4
3
2
1
0
1
happy
aggressive
calm/melancholic

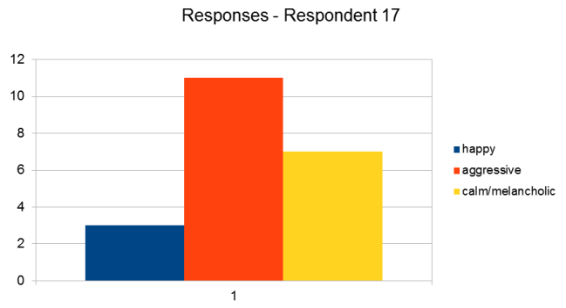

Responses - Respondent 12
10
9
8
7
6
5
4
3
2
1
0
1
happy
aggressive
calm/melancholic

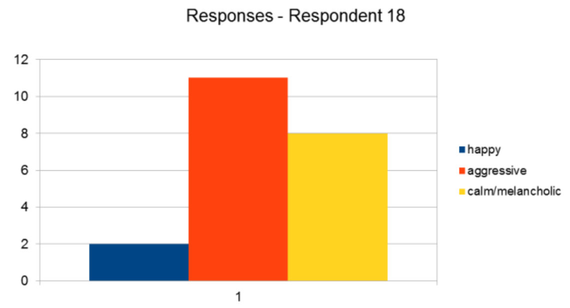

Responses - Respondent 13
12
10
8
6
4
2
0
1
happy
aggressive
calm/melancholic

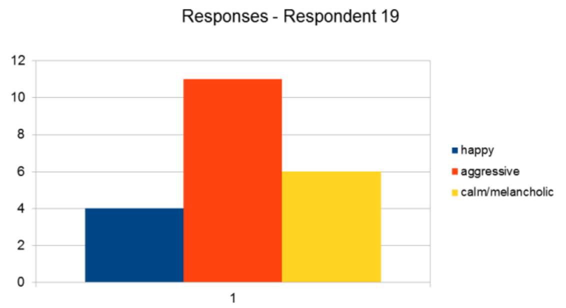

Responses - Respondent 14
14
12
10
8
6
4
2
0
1
Series1
Series2
Series3

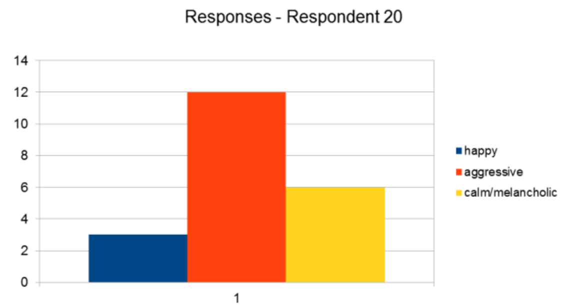

Responses - Respondent 15
12
10
8
6
4
2
0
1
happy
aggressive
calm/melancholic

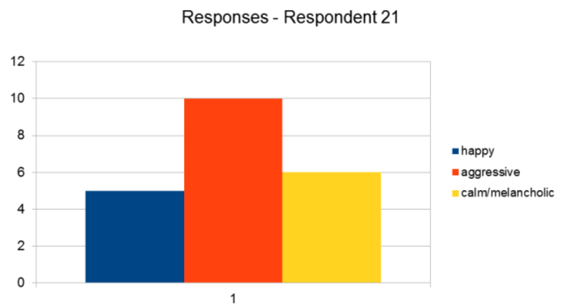

Responses - Respondent 16
12
10
8
6
4
2
0
1
happy
aggressive
calm/melancholic

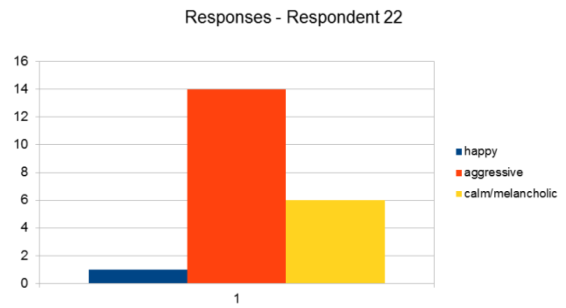

Responses - Respondent 17
12
10
8
6
4
2
0
1
happy
aggressive
calm/melancholic

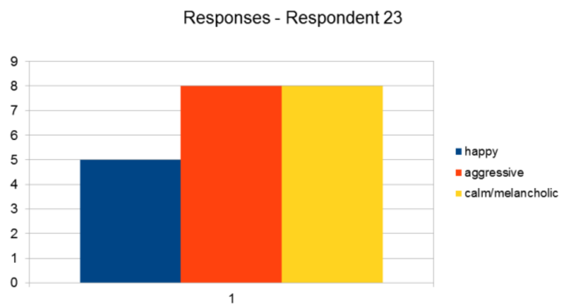

Responses - Respondent 18
12
10
8
6
4
2
0
1
happy
aggressive
calm/melancholic

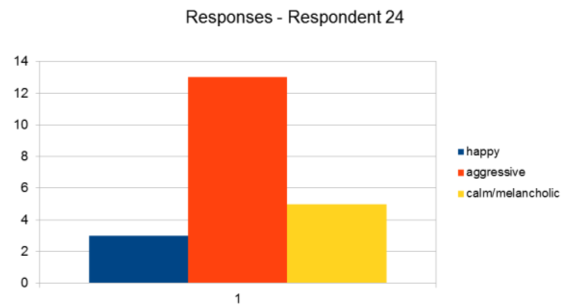

Responses - Respondent 19
12
10
8
6
4
2
0
1
happy
aggressive
calm/melancholic

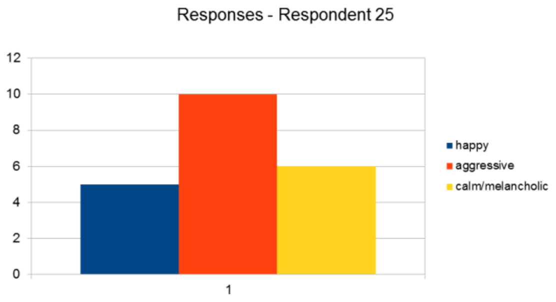

Responses - Respondent 20
14
12
10
8
6
4
2
0
1
happy
aggressive
calm/melancholic

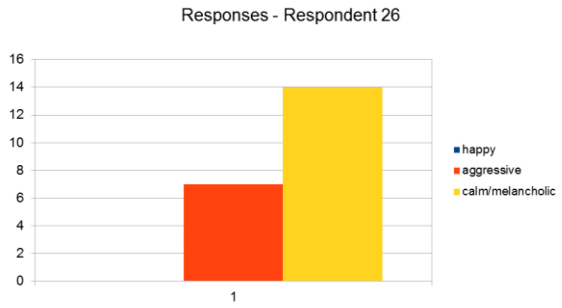

Responses - Respondent 21
12
10
8
6
4
2
0
1
happy
aggressive
calm/melancholic

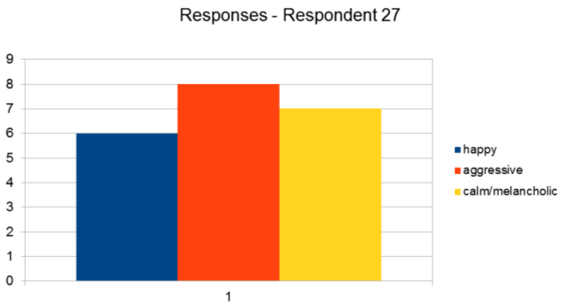

Responses - Respondent 22
16
14
12
10
8
6
4
2
0
1
happy
aggressive
calm/melancholic

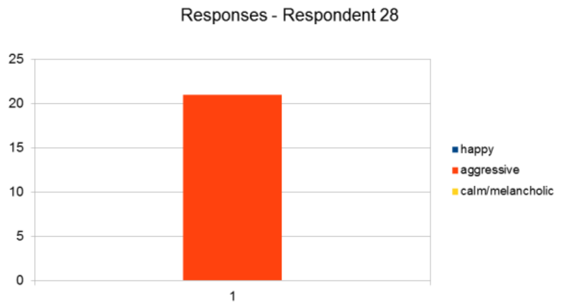

Responses - Respondent 23
9
8
7
6
5
4
3
2
1
0
1
happy
aggressive
calm/melancholic

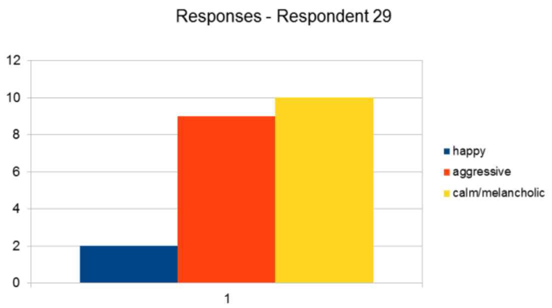

Responses - Respondent 24
14
12
10
8
6
4
2
0
1
happy
aggressive
calm/melancholic

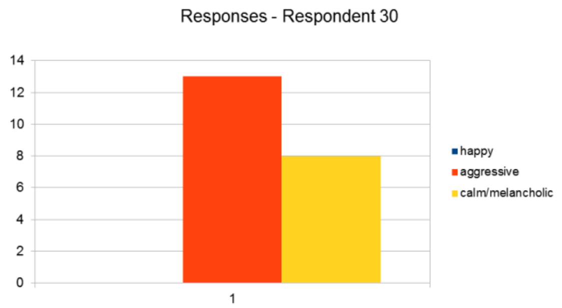

Responses - Respondent 25
12
10
8
6
4
2
0
1
happy
aggressive
calm/melancholic

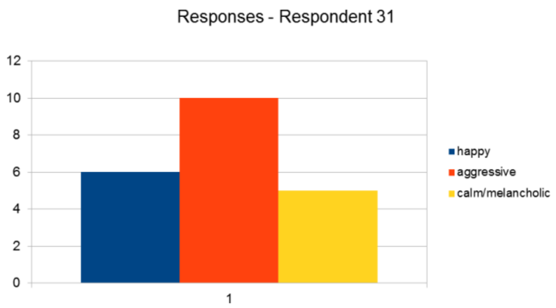

Responses - Respondent 26
16
14
12
10
8
6
4
2
0
1
happy
aggressive
calm/melancholic


Responses - Respondent 27
9
8
7
6
5
4
3
2
1
0
1
happy
aggressive
calm/melancholic

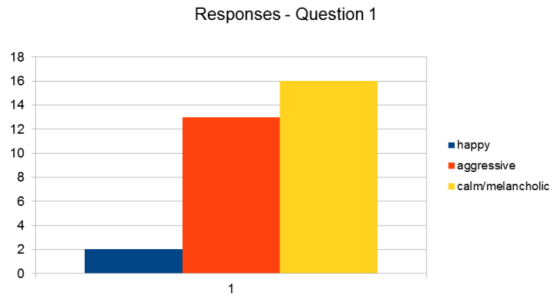

Responses - Respondent 28
25
20
15
10
5
0
1
happy
aggressive
calm/melancholic

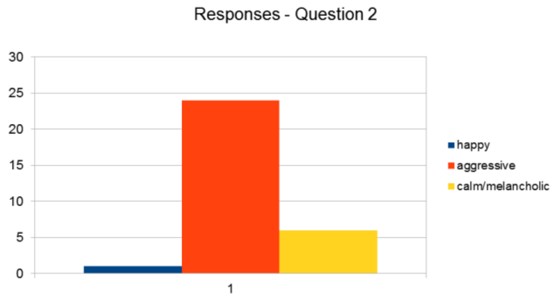

Responses - Respondent 29
12
10
8
6
4
2
0
1
happy
aggressive
calm/melancholic

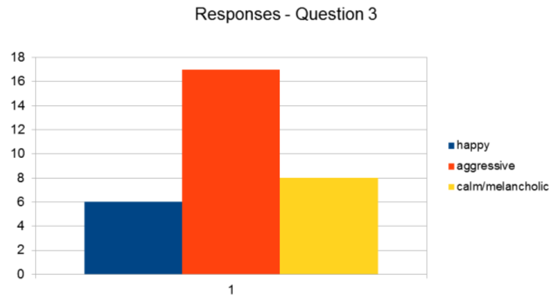

Responses - Respondent 30
14
12
10
8
6
4
2
0
1
happy
aggressive
calm/melancholic

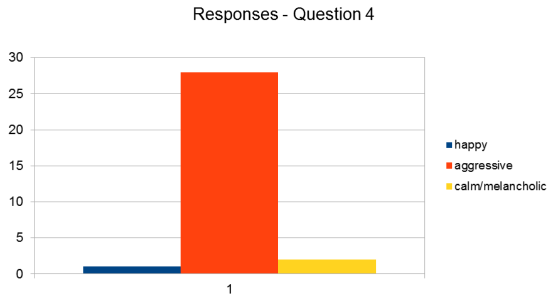

Responses - Respondent 31
12
10
8
6
4
2
0
1
happy
aggressive
calm/melancholic

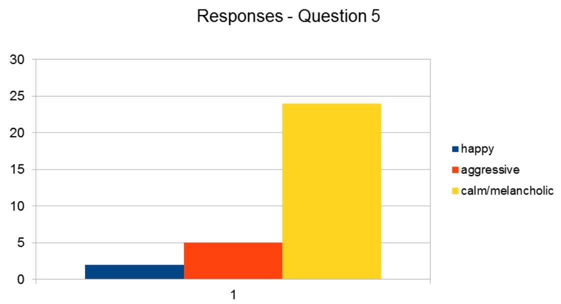
9.8.2 Responses per question for all respondents

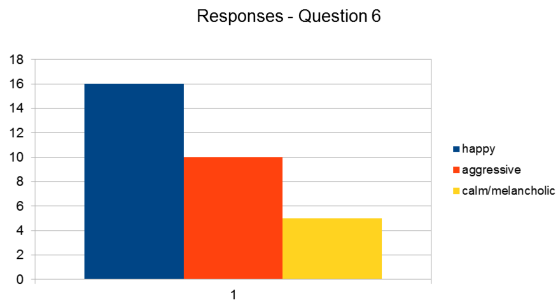


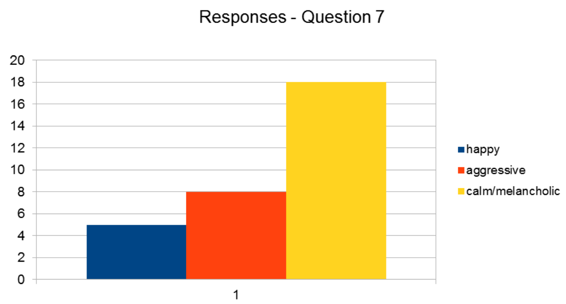


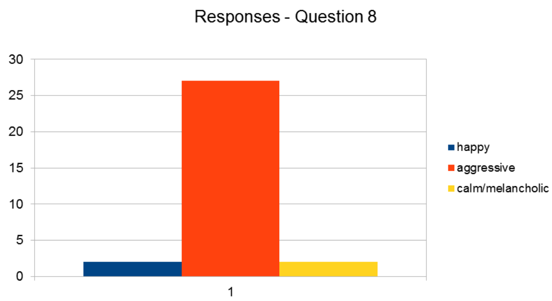


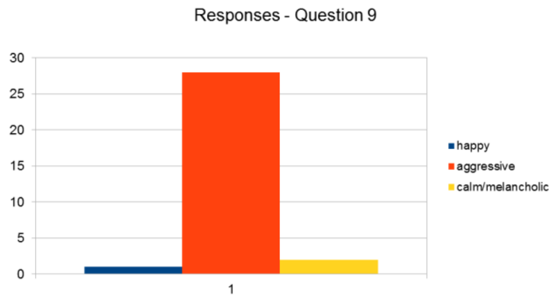


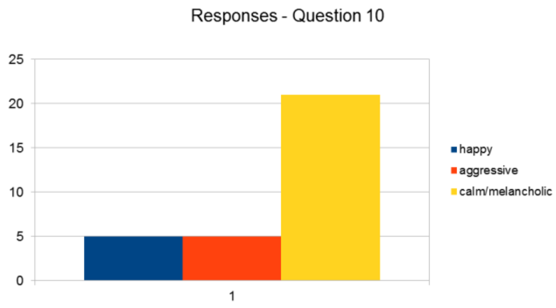


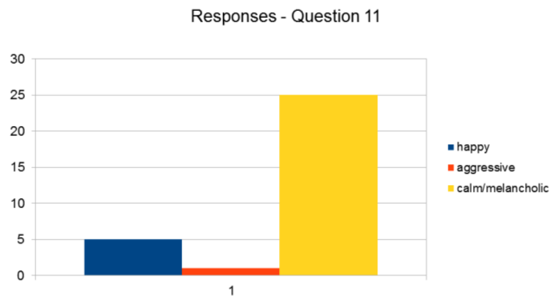


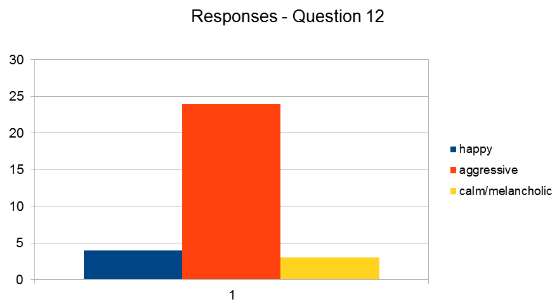


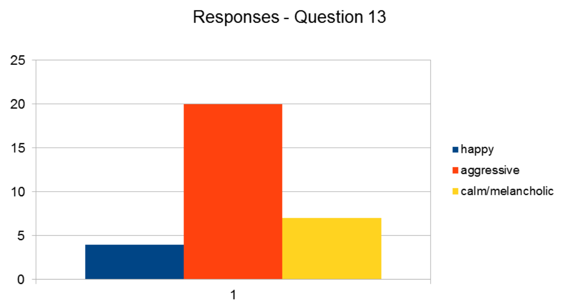


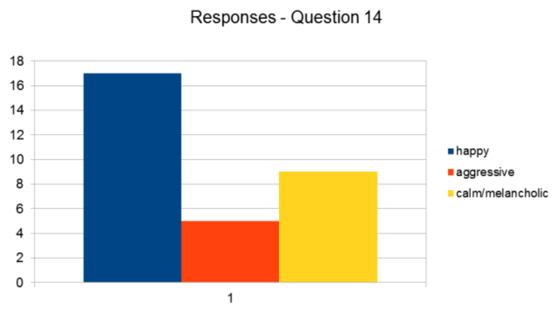

Responses - Question 9
30
25
20
15
10
5
0
1
happy
aggressive
calm/melancholic

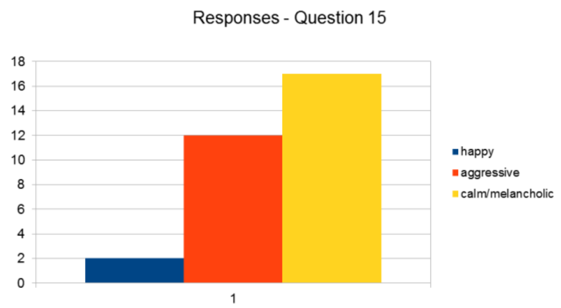

Responses - Question 10
25
20
15
10
5
0
1
happy
aggressive
calm/melancholic

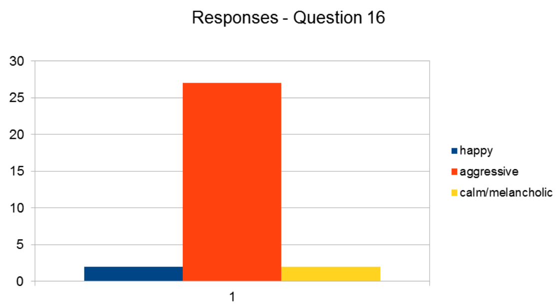

Responses - Question 11
30
25
20
15
10
5
0
1
happy
aggressive
calm/melancholic

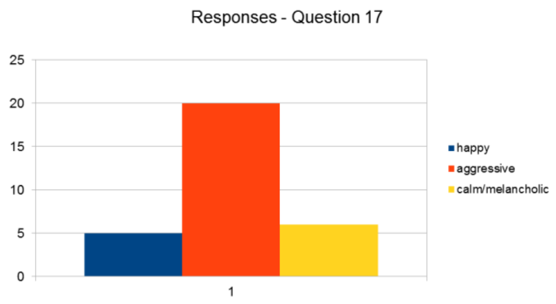

Responses - Question 12
30
25
20
15
10
5
0
1
happy
aggressive
calm/melancholic

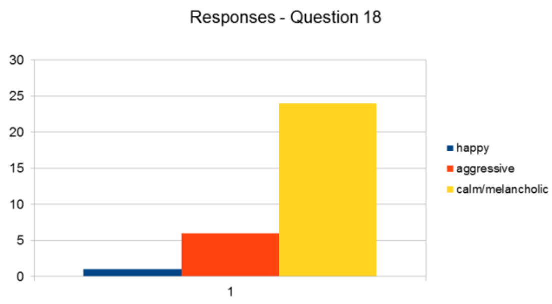

Responses - Question 13
25
20
15
10
5
0
1
happy
aggressive
calm/melancholic

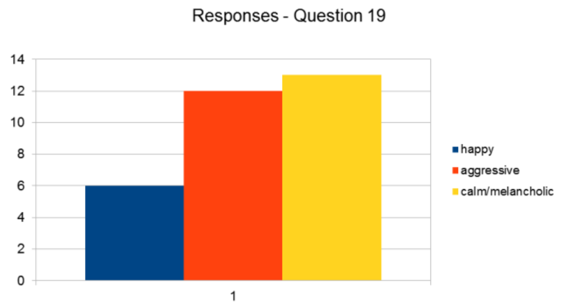

Responses - Question 14
18
16
14
12
10
8
6
4
2
0
1
happy
aggressive
calm/melancholic

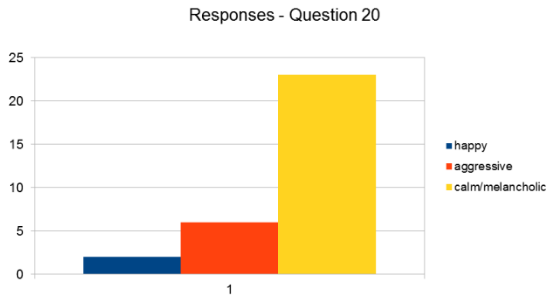

Responses - Question 15
18
16
14
12
10
8
6
4
2
0
1
happy
aggressive
calm/melancholic

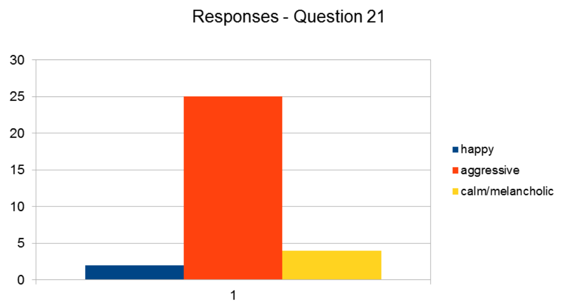

Responses - Question 16
30
25
20
15
10
5
0
1
happy
aggressive
calm/melancholic


Responses - Question 17
25
20
15
10
5
0
1
happy
aggressive
calm/melancholic


Responses - Question 18
30
25
20
15
10
5
0
1
happy
aggressive
calm/melancholic


Responses - Question 19
14
12
10
8
6
4
2
0
1
happy
aggressive
calm/melancholic


Responses - Question 20
25
20
15
10
5
0
1
happy
aggressive
calm/melancholic


Responses - Question 21
30
25
20
15
10
5
0
1
happy
aggressive
calm/melancholic

9.9 APPENDIX I - RAW DATA

Respondent	Age	Question 01	Question 02	Question 03	Question 04	Question 05	Question 06
		Response	Response	Response	Response	Response	Response
1	50	aggressive	aggressive	aggressive	aggressive	calm/melancholic	happy
2	36	calm/melancholic	aggressive	aggressive	aggressive	aggressive	calm/melancholic
3	56	calm/melancholic	calm/melancholic	aggressive	aggressive	calm/melancholic	aggressive
4	28	aggressive	aggressive	aggressive	aggressive	calm/melancholic	aggressive
5	27	aggressive	aggressive	aggressive	aggressive	aggressive	calm/melancholic
6	35	calm/melancholic	aggressive	calm/melancholic	aggressive	calm/melancholic	happy
7	51	calm/melancholic	aggressive	calm/melancholic	aggressive	calm/melancholic	happy
8	55	happy	aggressive	happy	calm/melancholic	happy	aggressive
9	23	calm/melancholic	calm/melancholic	calm/melancholic	aggressive	calm/melancholic	calm/melancholic
10	25	aggressive	calm/melancholic	aggressive	aggressive	calm/melancholic	aggressive
11	57	calm/melancholic	aggressive	aggressive	aggressive	aggressive	happy
12	30	calm/melancholic	aggressive	happy	aggressive	calm/melancholic	happy
13	46	aggressive	aggressive	calm/melancholic	aggressive	calm/melancholic	happy
14	55	aggressive	calm/melancholic	aggressive	aggressive	calm/melancholic	aggressive
15	25	calm/melancholic	aggressive	happy	aggressive	calm/melancholic	aggressive
16	67	aggressive	calm/melancholic	happy	calm/melancholic	calm/melancholic	happy
17	56	aggressive	aggressive	aggressive	aggressive	calm/melancholic	happy
18	57	calm/melancholic	aggressive	aggressive	aggressive	calm/melancholic	aggressive
19	26	aggressive	aggressive	calm/melancholic	aggressive	calm/melancholic	happy
20	43	calm/melancholic	aggressive	happy	aggressive	calm/melancholic	happy
21	38	calm/melancholic	aggressive	aggressive	aggressive	calm/melancholic	happy
22	28	aggressive	aggressive	aggressive	aggressive	calm/melancholic	happy
23	24	calm/melancholic	aggressive	aggressive	aggressive	calm/melancholic	happy
24	36	aggressive	calm/melancholic	happy	aggressive	calm/melancholic	happy
25	40	calm/melancholic	happy	calm/melancholic	aggressive	aggressive	happy
26	32	calm/melancholic	aggressive	calm/melancholic	aggressive	calm/melancholic	calm/melancholic
27	39	calm/melancholic	aggressive	aggressive	happy	calm/melancholic	aggressive
28	61	aggressive	aggressive	aggressive	aggressive	aggressive	aggressive
29	28	calm/melancholic	aggressive	calm/melancholic	aggressive	calm/melancholic	calm/melancholic
30	28	aggressive	aggressive	aggressive	aggressive	calm/melancholic	aggressive
31	26	happy	aggressive	aggressive	aggressive	happy	happy
	happy	2	1	6	1	2	16
	aggressive	13	24	17	28	5	10
	calm/melancholic	16	6	8	2	24	5
	total	31	31	31	31	31	31
	% happy	6.5%	3.2%	19.4%	3.2%	6.5%	51.6%
	% aggressive	41.9%	77.4%	54.8%	90.3%	16.1%	32.3%
	% calm/melancholic	51.6%	19.4%	25.8%	6.5%	77.4%	16.1%

Question 07	Question 08	Question 09	Question 10	Question 11	Question 12	Question 13	Question 14
Response	Response	Response	Response	Response	Response	Response	Response
calm/melancholic	aggressive	aggressive	calm/melancholic	calm/melancholic	aggressive	happy	happy
calm/melancholic	aggressive	aggressive	calm/melancholic	calm/melancholic	aggressive	aggressive	calm/melancholic
aggressive	aggressive	aggressive	aggressive	calm/melancholic	calm/melancholic	aggressive	aggressive
calm/melancholic	aggressive	calm/melancholic	calm/melancholic	calm/melancholic	aggressive	aggressive	calm/melancholic
calm/melancholic	aggressive	aggressive	calm/melancholic	calm/melancholic	aggressive	calm/melancholic	happy
calm/melancholic	aggressive	aggressive	happy	happy	aggressive	aggressive	happy
calm/melancholic	aggressive	aggressive	calm/melancholic	calm/melancholic	aggressive	happy	calm/melancholic
happy	happy	aggressive	calm/melancholic	calm/melancholic	happy	aggressive	happy
aggressive	aggressive	aggressive	calm/melancholic	calm/melancholic	aggressive	aggressive	calm/melancholic
calm/melancholic	happy	aggressive	calm/melancholic	calm/melancholic	aggressive	happy	aggressive
happy	calm/melancholic	calm/melancholic	aggressive	happy	happy	aggressive	calm/melancholic
calm/melancholic	aggressive	aggressive	calm/melancholic	calm/melancholic	happy	calm/melancholic	happy
calm/melancholic	calm/melancholic	aggressive	happy	calm/melancholic	aggressive	aggressive	happy
aggressive	aggressive	aggressive	calm/melancholic	calm/melancholic	aggressive	aggressive	aggressive
calm/melancholic	aggressive	aggressive	aggressive	calm/melancholic	aggressive	calm/melancholic	happy
calm/melancholic	aggressive	aggressive	calm/melancholic	calm/melancholic	aggressive	calm/melancholic	happy
calm/melancholic	aggressive	aggressive	calm/melancholic	calm/melancholic	aggressive	aggressive	happy
calm/melancholic	aggressive	aggressive	calm/melancholic	happy	aggressive	aggressive	happy
calm/melancholic	aggressive	aggressive	happy	calm/melancholic	aggressive	aggressive	happy
aggressive	aggressive	aggressive	calm/melancholic	calm/melancholic	aggressive	calm/melancholic	happy
happy	aggressive	aggressive	happy	happy	calm/melancholic	aggressive	happy
calm/melancholic	aggressive	aggressive	calm/melancholic	calm/melancholic	aggressive	aggressive	aggressive
aggressive	aggressive	aggressive	calm/melancholic	calm/melancholic	calm/melancholic	happy	happy
aggressive	aggressive	aggressive	calm/melancholic	calm/melancholic	aggressive	aggressive	happy
happy	aggressive	aggressive	happy	calm/melancholic	aggressive	calm/melancholic	happy
calm/melancholic	aggressive	aggressive	calm/melancholic	calm/melancholic	aggressive	calm/melancholic	calm/melancholic
happy	aggressive	aggressive	calm/melancholic	calm/melancholic	aggressive	aggressive	happy
aggressive	aggressive	aggressive	aggressive	aggressive	aggressive	aggressive	aggressive
aggressive	aggressive	aggressive	calm/melancholic	happy	aggressive	aggressive	calm/melancholic
calm/melancholic	aggressive	aggressive	calm/melancholic	calm/melancholic	aggressive	aggressive	calm/melancholic
calm/melancholic	aggressive	happy	aggressive	calm/melancholic	happy	aggressive	calm/melancholic
5	2	1	5	5	4	4	17
8	27	28	5	1	24	20	5
18	2	2	21	25	3	7	9
31	31	31	31	31	31	31	31
16.1%	6.5%	3.2%	16.1%	16.1%	12.9%	12.9%	54.8%
25.8%	87.1%	90.3%	16.1%	3.2%	77.4%	64.5%	16.1%
58.1%	6.5%	6.5%	67.7%	80.6%	9.7%	22.6%	29.0%

Question 15	Question 16	Question 17	Question 18	Question 19	Question 20	Question 21
Response	Response	Response	Response	Response	Response	Response
calm/melancholic	aggressive	aggressive	calm/melancholic	happy	calm/melancholic	aggressive
calm/melancholic	aggressive	aggressive	aggressive	aggressive	calm/melancholic	aggressive
aggressive	aggressive	aggressive	calm/melancholic	aggressive	calm/melancholic	aggressive
happy	aggressive	aggressive	calm/melancholic	calm/melancholic	calm/melancholic	calm/melancholic
aggressive	aggressive	calm/melancholic	calm/melancholic	calm/melancholic	calm/melancholic	aggressive
calm/melancholic	happy	aggressive	calm/melancholic	calm/melancholic	calm/melancholic	happy
calm/melancholic	aggressive	happy	aggressive	calm/melancholic	happy	aggressive
calm/melancholic	aggressive	calm/melancholic	aggressive	calm/melancholic	calm/melancholic	aggressive
aggressive	aggressive	aggressive	calm/melancholic	aggressive	calm/melancholic	aggressive
aggressive	aggressive	calm/melancholic	calm/melancholic	calm/melancholic	happy	aggressive
happy	calm/melancholic	happy	aggressive	calm/melancholic	aggressive	calm/melancholic
calm/melancholic	aggressive	aggressive	calm/melancholic	happy	calm/melancholic	aggressive
calm/melancholic	aggressive	aggressive	calm/melancholic	aggressive	calm/melancholic	aggressive
calm/melancholic	aggressive	calm/melancholic	calm/melancholic	calm/melancholic	calm/melancholic	aggressive
calm/melancholic	aggressive	aggressive	calm/melancholic	calm/melancholic	calm/melancholic	aggressive
calm/melancholic	aggressive	calm/melancholic	calm/melancholic	happy	aggressive	aggressive
calm/melancholic	aggressive	aggressive	calm/melancholic	aggressive	calm/melancholic	happy
calm/melancholic	aggressive	aggressive	calm/melancholic	calm/melancholic	calm/melancholic	aggressive
calm/melancholic	aggressive	aggressive	happy	aggressive	calm/melancholic	aggressive
aggressive	aggressive	aggressive	aggressive	calm/melancholic	aggressive	aggressive
calm/melancholic	aggressive	aggressive	calm/melancholic	aggressive	calm/melancholic	aggressive
aggressive	aggressive	aggressive	calm/melancholic	calm/melancholic	aggressive	aggressive
calm/melancholic	aggressive	happy	calm/melancholic	happy	calm/melancholic	aggressive
aggressive	aggressive	aggressive	calm/melancholic	aggressive	aggressive	aggressive
aggressive	aggressive	aggressive	calm/melancholic	aggressive	calm/melancholic	aggressive
calm/melancholic	aggressive	calm/melancholic	calm/melancholic	calm/melancholic	calm/melancholic	aggressive
aggressive	happy	happy	calm/melancholic	happy	calm/melancholic	calm/melancholic
aggressive	aggressive	aggressive	aggressive	aggressive	aggressive	aggressive
aggressive	calm/melancholic	aggressive	calm/melancholic	happy	calm/melancholic	calm/melancholic
calm/melancholic	aggressive	aggressive	calm/melancholic	aggressive	calm/melancholic	aggressive
aggressive	aggressive	happy	calm/melancholic	aggressive	calm/melancholic	aggressive
2	2	5	1	6	2	2
12	27	20	6	12	6	25
17	2	6	24	13	23	4
31	31	31	31	31	31	31
6.5%	6.5%	16.1%	3.2%	19.4%	6.5%	6.5%
38.7%	87.1%	64.5%	19.4%	38.7%	19.4%	80.6%
54.8%	6.5%	19.4%	77.4%	41.9%	74.2%	12.9%

9.10 APPENDIX J - LISTENING MATERIALS

See Compact Disc attached