Composing Speech:

Investigation and Application of Musical Expression Embedded in Spoken Language

Marc Du Plessis

1043671

Dissertation submitted in partial fulfilment of the requirements
for the degree of Master of Music
in the Wits School of Arts, Faculty of Humanities,
University of the Witwatersrand

Johannesburg, March 2024

Acknowledgments

This dissertation has been a challenging and rewarding learning experience which I could not have completed without the help of my two supervisors, Dr Cameron Harris and Dr Jonathan Crossley. Their constructive critique has equipped me with the skills to write this dissertation and compose the supporting music in a manner that leaves me feeling proud and accomplished.

I would like to say a special word of thanks to my partner Christy, my parents, Charl and Nicole, and my friends who have supported me in my passion and given me the help and guidance needed.

Declaration

I declare that this research thesis is my own unaided work. It is submitted for the degree of Master of Music at the University of the Witwatersrand, Johannesburg. It has not been submitted before for any other degree or examination at any other university.

Marc Du Plessis

12 March 2024

Abstract

This dissertation explores the musical potential of emotive expression in cut-up speech sounds. Cut-up is a twentieth-century technique with roots in Dadaism in which one cuts "pre-existing material into radical juxtapositions" (BBC, 2015), made popular in literature by William Burroughs in the 1950s and 60s. Speech is used primarily to communicate information relating to the world around us, but it operates sonically. Therefore, it has inherent parameters that can be manipulated to inform how information is received. The ability to manipulate the inherent sonic parameters of speech is one way in which it can be emotively coded. Sung vocals with lyrical content in music differ from speech in that the roles of information communication and the manipulation of the sonic parameters are reversed. Where speech relies on the manipulation of sonic parameters to augment or diminish the information being conveyed, sung vocals that utilise lyrical content rely on the semantic content to augment or diminish the sonic characteristics of the voice. Sung vocals could therefore be thought of as sonic utterances that are semantically coded. These inherent parameters are shared by music as it also operates in the sonic realm. The researcher used electronic music production techniques to isolate the shared parameters between music and speech (pitch, rhythm, timbre, and dynamics), and composed expressive, accessible, and engaging musical works based on these parameters. Digital music technology has the capacity to explore the limitations of sonic expression, due to its capacity to manipulate recorded sound waves. Therefore, it equipped the researcher with the necessary tools to manipulate cut-up speech sounds with compositional intent.

The objective of this research was to compose musical works that drew from popular music styles, with an aesthetic focus on rich, timbrally expressive vocal material created from recordings of speech, to understand the expressive capabilities of the chosen raw material (speech sounds). The methodological procedure was to record speech from various sources, edit (cut-up) the phrases to create brief clips that were divorced from semantic signification, present the edited clips to an audience, and analyse their responses. The researcher used the insights from this analytical process to inform the use of the same speech sounds in the compositional practice. The researcher presented 26 examples (brief composed cut-ups of speech sounds) to 45 participants in a survey group and eight South African music industry professionals in one-on-one interviews. The responses yielded qualitative data that was analysed using thematic coding, followed by statistical analysis using Spearman's rank

correlation (1904). The results provided vague answers to the primary research questions, but ultimately supplied the researcher with various qualitative interpretations of how the speech sounds expressed meaning in a cut-up context. This informed the researcher's creative practice in the musical application of cut-up speech.

Although the interpretation of the qualitative data did not result in definitive answers to the research questions, the aim of this research to explore the musical application of emotive expression in speech was achieved. The understanding that a listener experiences music in an inter-subjective and inter-contextual manner, combined with the expressive nature of the raw materials, liberated the researcher to compose expressive music without the need to know each listener's subjective experience of expression.

Musical works:

https://drive.google.com/drive/folders/1gdVt2qVFIa4I0CnfkoWeOAVBEylDqR-h?usp=drive_link

Unprocessed Examples (Original Sound Edits):

https://drive.google.com/drive/folders/1Z48UKr37gaqX0TIiaxuhYVOGr5wWEshF?usp=drive_link

Processed Examples (Survey and Interview Edits):

https://drive.google.com/drive/folders/1_Jqi58y12FVrqOixsQQZpj1XqIaApYZP?usp=drive_link

Ethics clearance number:

H21/10/07

Table of Contents

Index of Tables and Figures

Tables

Chapter 3

Chapter 4

Chapter 6

Figures

Chapter 2

Chapter 6

1. Introduction

1.1 Aim

"Now listen to this." The words were smudged together. They snarled and whined and barked. It was as if the words themselves were called in question and forced to give up their hidden meanings (Burroughs 1962: 21).

Speech and music share expressive parameters such as pitch, rhythm, timbre, and dynamics. This research explores three primary questions to determine if one's ability to interpret emotive meaning in both mediums is directly linked to this phenomenon. These questions were explored through utilising a mixed methodology of qualitative and quantitative analysis, and through the creative act of composition. Through the latter, the researcher explored the extent to which the expressive parameters of speech can be utilised to create musical material. By divorcing speech sounds from their original syntactic structure and utilising them in the creation of musical works, the researcher composed music that is primarily emotive, but also accessible in-so-far-as a general listener may be able to engage with, and interpret meaning from, the music. The music is presented in the form of an eleven-track electronic album which borrows aesthetically and structurally from popular music. The musical works were inspired by Salman Rushdie's *Midnight's Children* (1981), in which the protagonist is born with the gift to hear the thoughts his fellow characters. As such, each track in the album embodies a metaphysical representation of fictional characters imagined by the researcher.

This research interrogates three research questions in relation to the emotive expression of English speech sounds in a cut-up form.

(1) Do speech and music share parameters that express emotive meaning in an analogous manner?

Both music and speech are sonic-based mediums of communication and make use of pitch, rhythm, timbre, and dynamics: is our emotive experience of both mediums analogous? Literature pertaining to the relationship between language and music will be reviewed and used to analyse the qualitative results of this question discussed in Chapter 4.

(2) How does the act of composing with cut-up speech sounds change the emotive meaning for the receiver?

This research aims to investigate whether the emotive signification of speech is altered by rearranging its syntax based on formal musical structures. In order to derive data-driven insight into this question, the researcher created a listening experiment that was presented to two groups of participants. One comprised university music students and non-music specialist members of the public, and the other involved one-on-one interviews with South African music industry experts. This question was also considered by the researcher in his own reflection on his musical works. The technique of rearranging speech sounds was inspired by William Burroughs, an American writer who helped popularise the styles of writing and literature known as "cut-up" in *The Nova Trilogy* (commonly referred to as *The Cut-up Trilogy*), a series of three novels written between 1961 and 1964. The novels were constructed:

> using the 'cut-up' method in which existing texts, including Burroughs' own writing and/or writings by other authors, were physically cut into pieces of variable length and re-assembled in random order to generate unexpected juxtapositions and new syntactic relationships (Murphy 2002).

The researcher was initially introduced to Burroughs' work through the song "Fire is Coming" by Flying Lotus featuring David Lynch (2019). The song features cut-up-style lyrics narrated in a Burroughs-esque voice by David Lynch. Burroughs' writing subsequently became a source of inspiration as the technique and resulting experience of the writing style felt analogous to the practice of composition. Like Burroughs, the researcher cut existing material (recordings of speech) and reassembled the pieces to create new syntactic relationships between the sonic artifacts. The researcher's approach differed from Burroughs' in that the cut-ups were created from a single phrase of recorded speech, by removing utterances from the original phrases. This process resulted in new syntactic relationships between the remaining speech utterances. These cut-ups became the raw materials used for the album tracks. Burroughs' application of the cut-up technique, on the other hand, joined strings of words, sentences, and occasionally passages, with those from other sections or source material.

(3) Can a composer extract and harness the speech sounds which allow one to signify emotive meaning through spoken language?

This question investigates the problem of signification. Is it possible to isolate and harness the speech sounds that signify emotive expression from the audio of recorded speech? By

divorcing linguistic meaning from speech, is the emotive component of spoken language more easily identifiable? One of the methods for investigating this question was to create a series of 26 brief audio examples that exhibited the cut-up method and present these examples to participants, both in a group survey and one-on-one interviews. The data gathered would be analysed using qualitative thematic coding as described by Virginia Braun and Victoria Clarke in *Successful Qualitative Research* (2013) and quantitative data analysis using Spearman's Rank Correlation Coefficient to explore potential connections between participants' responses.

This dissertation and its accompanying album should be viewed as the culmination of process-based artistic research (research through art and design (Frayling 1993: 5), with each operating iteratively towards the same end. Namely, to investigate the extent to which music and speech communicate emotive meaning in an analogous manner. The nature of the relationship between language and music has been considered at length for centuries. Scholar Michael Davis discusses Jean Jacques Rousseau's (1712-1778) interpretation of the relationship between language and music in his paper "The Music of Reason in Rousseau's 'Essay on the Origin of Languages'" (2012). In this paper Davis explains that Rousseau's approach stems from the notion that language developed because humans experience one another as subjects rather than objects (2012: 392). Subjects require a logical system (language) through which they can interact while objects are able to rely on sensations communicated through imitative expressions such as painting or music (Davis 2012: 392-393). In this interpretation humans are subjects that can perceive themselves in the world, and therefore require methods of communication (language) that allow them to interact with other humans about their experiences and the external world around them. In Rousseau's view (as described by Davis) objects include animals and rely on instinct and feeling to navigate the world, sensations that are imitated in artistic works.

This dissertation does not conform to Rousseau's interpretation of the relationship between music and language. Rousseau implies a binary separation between musical, or artistic, expression and language, presenting a problematic perspective of people who are, in Rousseau's view, capable of subjective and objective experiences. Instead, the researcher has chosen to approach the relationship of music and language as two necessary modes of communication that are connected through their nature as temporal sonic mediums and work together in both forms to convey information about the human experience.

This project emerged from a desire to understand which elements of spoken language are analogous to modes of musical expression and explore the ways in which the emotive content of spoken language could be used in a musical context. Musical and linguistic expression are continuously combined in popular music, and the researcher wished to investigate this phenomenon by examining the role of the emotive delivery of cut-up speech sounds and how they can impact the overall expression of a musical work.

The album composed in conjunction with this dissertation makes use of instrumentation limited to cut-up speech sounds and drum kit. The drums are the musical instrument through which this researcher artistically expresses himself. The voice was of compositional interest because it is the primary instrument through which most people communicate with one another. Our conceptual lives, and consequently, our perceptual lives, are born, grow, and develop through our experience with language (Langer 1942: 126). As a result, one can communicate and cultivate relationships with others. The combination of these two elements appeared to suggest a compositional framework through which to explore conceptual and creative expression.

The approach of incorporating live drums and speech in the context of pop-influenced electronic music has been previously explored by American drummer, Zack Danziger. Danziger used hybrid acoustic and electronic drum kits to perform in Wednesday Night Titans, a duo comprising Danziger and bassist Kevin Scott. Their music combines 1980s wrestling videos, a hybrid drum kit setup of live drums and electronic triggers and sample pads, with live bass. The result is a multimedia performance that allows them to improvise with the audio (speech) and visuals of the wrestling clips using both acoustic and electronic techniques in a live context.

1.2 Rationale

The researcher proposes that without intent and an understanding that the tools at our disposal are a means to an end, we give the tool the agency in our work. A tool carries out a particular function and when used without creative intent, will continue to produce a predictable outcome. Electronic music technology is no different. In the world, technology and music have a symbiotic relationship. Since the invention of the first musical instruments, technology has been used to create new sounds and develop new instrumentation to satisfy artistic conceptions. The documentation of music through music notation, the enhanced quality of the sound

produced during a performance through architecture (controlling the acoustic qualities of a space), and the ability to record music, are all examples of technological developments that allowed music to be experienced in different ways. These developments simultaneously laid the groundwork for new conceptions of music. Our interaction with technology as it relates to music production and distribution can be viewed as dialectic because our exploitation of the available technology is often restricted to the functionality of whatever tool we happen to be using. The common-place use of musical technology could imply some acceptance of the restrictions of the tools at our disposal. However, these restrictions are not fixed, as artistic and technological experimentation is continually redefining the boundaries of what is possible, and what is accepted.

The ability to record sounds for later playback has the innate ability to be changed by the medium through which it is heard. For example, magnetic tape allows for sounds to be played slower or faster depending on the speed setting of the tape machine. Vinyl record players have a similar ability. They can be played at standardised speeds (35, 45, and 75 revolutions per minute) or physically manipulated to alter the playback speed of the record. This concept highlights a key artistic consideration when exploring and creating electronic music. There needs to be artistic intent governing how technology is used to create art, otherwise the result could lack the depth required to be truly meaningful. In its current guise electronic technology has expedited the creation, distribution, and consumption of music. The researcher's decision to compose a popular music album using electronic music techniques creatively challenges the accepted restrictions of the medium. The adoption of formal aspects such as structure and instrumentation was achieved with an emphasis on presenting thematic material in an intentional and meaningful way.

Communication, how humans share ideas and information, is at the forefront of this dissertation's rationale, along with the desire to explore how communication adapts in relation to its medium. Language and speech were the natural next step, while the medium of electronic music allowed for flexibility and freedom in the exploration of compositional methods that would not limit the creative output. When we communicate with one another we are engaging in the transfer of information. This information can take the form of conceptual ideas, data, artistic expressions, or emotive expressions (amongst others) and by participating in the exchange of information we develop our understanding of the people and world around us.

Linear and Interactive Communication

There are two modes of communication called linear and interactive communication (Bock 2014: 36-37). Linear communication refers to a one-way sending of information to a receiver in which a sender encodes (speaks) a message to a receiver who then decodes the information and, hopefully, understands what is being communicated (36). In interactive communication we are constantly aware of the receiver and adjust the encoded messages to communicate in a particular environment or with a particular person more effectively. (37). Additionally, there is a semiotic element to the way we communicate. Languages are a system of signs that signify depending on the context in which we use them, but it is important to note that "signs are made and remade not used and sent" (45). This phenomenon happens as the result of framing: "the mental filters we use to interpret an interaction or text" (41) - and interpretation (45). This phenomenon can be applied to our engagement with music. On the surface, our interaction with music seems to be a linear form of communication: we are presented with an encoded message, and we decode and interpret meaning from what we hear. Just as language signs are made and remade depending on the framing or context, so too are musical signs. This concept greatly interests the researcher as a composer and influenced the creative aspect of this project. It is the researcher's goal to communicate emotive expression with the people that hear this project's music. The idea that through the reframing and the continuous remaking of signs, one can have novel and meaningful musical conversations, encourages the notion that a reframing of the "signs" of electronic music production can broaden the artistic scope of the medium.

Language can often fall short when trying to articulate the intricate nature of our sentient life, as our emotions can be complex and abstract, and not easily put into words (Langer 1942: 48). On the other hand, music excels in representing a connotative, symbolic, and subjective experience due to the two phenomena's (music and sentient life) "similarity of logical form" (28). This proposes that both music and one's felt emotions convey meaning in the same way. The meaning of music lies in the receiver's subjective feeling, just as it does with one's emotional experience. Although music and language function differently in conveying meaning (Adorno 2002: 113), speech may be implementing some of the same parameters that characterise musical expression to convey emotions and connotative meaning. Imagine hearing someone say, "Oh no, that's terrible" in an earnest and caring tone compared to a sarcastic or apathetic tone. The sentence and person are the same, yet one may make you feel comforted and understood, while the other may make you feel unheard or ridiculed. This example highlights the effect of our articulation on communication. The researcher suggests that the

reason we can change the way we speak, and subsequently evoke a different response from the people we communicate with, is because we are able to control the musical parameters of our speech patterns (pitch, rhythm, timbre, and dynamics).

According to philosopher and musicologist Theodor Adorno, music does not form a system of signs that could facilitate a true abstraction of what is experienced (2002: 113). However, much of contemporary popular music includes lyrics, which function as a vehicle for conceptual and linguistic information. “Language is conception, and conception is the frame of perception” (Langer 1942: 126). The voice as an instrument of communication is not merely a vehicle for conveying a system of signs that signify conceptual ideas. Public speakers, educators, politicians, and radio and podcast presenters are not just able to convey information in an eloquent manner, they are able to captivate or evoke an emotional response from their audience. As a composer this phenomenon presented an opportunity to use cut-up speech sounds and the musical parameters of speech to affect emotive expression in the creation of musical works.

Historically, music and language have developed similarly by incorporating form in a disciplined manner (Langer 1942: 216). Music and language, spoken and written, both rely on form to eloquently express their “import” (content) (ibid.), but what happens when the form of language and its import obey the grammar and syntax of a specific musical expression? In the same way that grammar functions in language using rules to govern how words and sentences are constructed, music can also be understood as having rules that facilitate the construction of melodic and harmonic phrases and form dependent on the place, time, and community in which the music has been created. The compositions have been created as a way of exploring how the emotive components of cut-up speech sounds, largely divorced of semantic meaning, behave in a musical context.

2. Literature Review

This chapter presents the literature used by this research as it pertains to the three research questions. The questions are each multifaceted and could be approached in a variety of ways. As a composer whose aim is to create musical works informed by their research, these questions were regarded as avenues for creative practice. An investigation into the literature below has been used to inform analysis and creative practice. Thus, it has enabled the researcher to consider how one might address the research questions from a compositional perspective.

The literature has been divided into the various foci that the researcher explores. Section 1 is concerned with musical accessibility, specifically outlining composer and writer Jochen Eisentraut (2013) and composer and scholar Dan Dediu's (2012) ideas that there are certain practices and qualities of accessible music that allow for easier engagement on the part of a listener. Section 2 is concerned with the relationship between language and music. This project makes use of speech sounds in its compositional practice and as a result, the idea that language and music share parameters has been explored in the literature. Section 3 is focused on the creative application of the cut-up method used by writer, William Burroughs. Section 4 deals with artistic research, a field that places emphasis on the production of knowledge through process-based creative practice. Section 5 discusses the expressive capabilities of popular music as it pertains to the aesthetic of the researcher's musical works. The creative output of this research takes the form of eleven electronic music tracks that conform to the format of a popular music album. Section 6 briefly introduces the qualitative and quantitative analysis methods that will be discussed in Chapter 3 as this project uses data gathered from respondents to inform its compositional output.

2.1 Accessibility in Music

This project centres the idea of accessibility around the aspects of a piece of music which a listener can recognise and to which they can relate. Those characteristics of a piece of music that sound or seem familiar to a listener. Composer and scholar Dan Dediu provides a definition for accessibility: "a phenomenon born of the relationship of communication between sender and receiver, by which the sent information has direct access to the mental space of the receiver of that information" (2012: 53).

Composer and writer Jochen Eisentraut created a typology for exploring accessibility in various musical situations. It is challenging to define a specific quality that music must possess to be received accessibly as "in considering musical accessibility we are looking at interactions between human beings and music (often via some media) and as such they involve an intricate interplay of musical structures, cultural constructs and psychology" (Eisentraut 2013: 11-12). For music to be accessible within the boundaries of Eisentraut's proposal, it not only needs to be physically accessible to people, but it also needs to interact with culture for human beings to be receptive to its meaning. The researcher is interested in how structures of music, timbral qualities in the sound sources used, and/or aesthetic choices made by the composer may allow the music created in this research to be received in a more accessible manner.

Eisentraut's typology divides musical accessibility into three levels: (1) contact, (2) reception, and (3) participation (2013: 15). The levels, although specific in their focus, are not to be thought of as isolated and self-contained, as every situation can be analysed from each perspective (Eisentraut 2013: 2). Contact describes how music needs to be experienced by a listener in a physical manner (21). If a listener never hears a piece of music, any discussion about accessibility is immediately halted. If someone wants to listen to music, they need ways to access it; these ways naturally change and evolve in conjunction with culture and technology, and they are influenced by socio-economic and geographical factors (21). Reception can be interpreted from a variety of perspectives: does the music communicate emotive meaning or accompany an activity in one's life and/or is one able to consciously interpret formal structures within the music that would facilitate a deeper or more complex understanding of what is being heard (22)? However, one does not need to be able to consciously interpret and evaluate musical form, or its technical attributes to receive input from music, because music can function as a regulator (relaxing music after a stressful experience), a confidant (music that helps one externalise their feelings), or a distraction from one's emotional experience (music that accompanies a mundane activity) (22).

Level 2 accessibility is a process of engagement that deals with aesthetics, individual psychology, and preference, and it is also subject to change in accordance with the lives of the individuals that experience it (Eisentraut 2013: 1, 9).

Level 3 accessibility deals with participation: “can I play an active part in that particular musical sphere?” (Eisentraut 2013: 15). Playing a musical instrument and learning to perform music falls into this level of accessibility. Socially interacting with a musical scene by dressing in a particular way or making friends with people who share the same musical taste also forms part of participation in music. “If we are keen enough on some musical genre, will we be able to become part of it in some way” (1). For a person to participate in some way with music or a musical social group, levels 1 and 2 are regarded as a prerequisite because without contact with the music or an affinity for it, it is unlikely any meaningful participation can take place (23). According to Eisentraut, if one engages in the level of participation, the potential for deeper understanding and appreciation is possible (45).

Eisentraut’s book, although providing a comprehensive examination of the accessibility of music, is more a study of the *species* of musical accessibility than it is a prescriptive guide to accessible music (Braae 2014: 395). His aim is to locate accessibility and make sense of its function within the discourse, but this dissertation wishes to focus on how one may use accessibility as a creative tool to communicate with a receiver. One critique of Eisentraut’s book is that he uses dualism to explain a concept that he states is pluralistic and multifaceted, which weakens his conceptual framework for accessibility. Anthropologist Evangelos Chrysagis suggests that the field would benefit more from an ethnographic approach to musical discourses (Chrysagis 2014: 369).

If a musician desires to consider the accessibility of their music, they risk creating a listening environment that does not present any new information (Dediu 2012: 50), in favour of a rearrangement of previously experienced musical phenomena. This leads to music that appears pleasant and confirms our established taste but does not allow us to gain any new perspectives apart from the potential of the lyrical content. Conversely, music that contains none of these accessible characteristics, although containing only new or novel material, has a potentially harder task of conveying expression to a larger audience (if that is the intention) due to a lack of grounding in an established musical environment.

If we consider music to be a medium of communication (Dediu 2012), then, like language, music must have certain recurring phrases, syntax or grammar that are both common and commonly understood. These require little-to-no effort on the part of recipient to decipher their meaning. Accessibility in music with the intention to convey an artistic message in a new or

novel way needs to consider the medium (firstly, music, and secondly, stylistic, or genre-based tropes) and the message (what is new and how does it augment the expectations of the listener?).

Dediu approaches the topic of accessibility in music from the perspective of a composer and musicologist. He explains the phenomenon of accessibility as both a relational phenomenon and something to be built. Accessibility is achieved in the act of discovery on the part of a receiver rather than something that a work of art inherently possesses (2012: 60). It is Dediu's belief that music encompasses five roles for the receiver: psychological, social, cultural, ideological, and moral. The idea of accessibility is crucial because, for the role to be successfully realised, understanding is required (49). On this point Dediu draws from Russian linguist I. M. Lotman and his concept of mental space, "the informational reservoir of a given individual at a certain moment in time, containing the totality of the information they hold" (50). For communication to occur between sender and receiver, there need to be common elements in both of their mental spaces. Without these, communication breaks down and no information is transferred. However, total overlap of elements in the mental space of the sender and receiver results in "communication without content" (50).

So, is this proposing that for an artistic work to be accessible it should have an *X* value of accessible content and *Y* value of new and novel information? Even if such a process was developed and used to create art, the outcome would most certainly fall short of "perfect" accessibility, in which every receiver was able to receive the desired new information. As Eisentraut states, accessibility operates on an individual level between receiver and artwork. Considering this, along with the understanding that cultural context is extremely variable, the scope of musical accessibility becomes evident. When addressing the idea that objects (in this case works of art) are accessible, Dediu has this to say: "an artistic object cannot actually be accessible, but only have in its structure elements that, in a certain context and in relation to a certain receiver, can act as catalysts for communication and help the emergence of the relational phenomenon of accessibility" (2012: 59).

2.2 Music and Language

It is beyond the scope of this research to detail a linguistic history and analysis of the various ways in which language functions to convey meaning, or to provide a definitive method for

understanding language. However, the musical source materials used in this project are cut-up speech sounds. Therefore, an investigation of language and music's ability to express meaning allowed the researcher to understand how these cut-up speech sounds could be utilised for composition. "Both speech and music are noted to be 'particulate' systems, in which a set of discrete elements of little inherent meaning are combined to form structure with a great diversity of meanings" (Dilley 2009: 535). This section will discuss philosophies surrounding musical expression and unpack linguistic definitions that have informed the research.

If one considers the notion that music is a system that can signify meaning (Dilley 2009: 535), then it is logical to ask: what kind of meaning does music signify? Suanne Langer, a mid-twentieth century philosopher believed that music conveys meaning because it can be understood as "a tonal analogue of emotive life" (Langer 1953: 27). In other words, music is experienced in the same way that one "feels" the reality of sentience. Music theorist, Raymond Monelle, reviewing Langer's conceptualisation of musical expression, states that, often complex and abstract, one's sentient experience is constantly in flux and can be challenging to conceptualise using language (Monelle 1992: 9). Music expresses, not in the sense that it directly conveys emotions experienced by the listener, but rather it is felt and understood as the "expression of an idea … it is a symbol by virtue of being felt as quality rather than recognised as a function" (8). Langer's "felt" idea is not limited to emotional information (e.g., happiness or sadness), rather, art can convey any feeling that one can experience, from our sense of touch and the feeling of pain or euphoria to abstract emotions and thoughts (Reece 1977: 45).

What, then, makes a piece of music, or any other work of art, successful? According to Langer, the task that every artist faces is to effectively, through symbolism, convey an individual and subjective experience objectively (Langer 1964: 381). Following this proposal, an artist would use the knowledge of their medium to create an artistic object that acts as a representation of their subjective experience. The artistic object, however, cannot converse with a receiver. Its meaning is encoded by the artist through their ability to manipulate the medium. Receivers interact with the artistic object from their own subjective perspective and interpret meaning based on their knowledge of the symbolism. The successful transmission of message between the artist and the receiver, as per Langer's interpretation, is then dependent on how well the transformation has been constructed.

The role of symbolism is to represent concepts and ideas (Langer 1953: 26). Signs refer to or suggest specific entities, the way that a noun functions in language. However, symbols can express ideas to a receiver, and often express a multitude of ideas to a multitude of receivers based on context and discourse (26). Therefore, one could choose to interpret the success of an artistic work in its ability to move beyond signification into the world of symbolism. This statement presents a pertinent discussion in relation to the researcher's third primary research question: Can a composer extract and harness the speech sounds which allow one to signify emotive meaning through spoken language? If the success of an artistic work requires it to live in the world of symbolism, inside which a multitude of ideas can be expressed to a multitude of receivers irrespective of cultural context or discourse, then supposing a composer would be able to extract that which signifies emotive meaning through language, would the music be a successful work of art?

Musicologist Simon Emmerson's insight speaks to Langer's idea that music can communicate a felt or experienced version of our sentient realities. Emmerson's idea of "image" presents the concept that the sounds used by a composer can evoke in the listener, an "image" (Emmerson 1986: 17) that is not simply the representation of the sonic stimulus, but an "image" or idea "lying somewhere between true synaesthesia with visual image and a more ambiguous complex of auditory, visual and emotional stimuli" (17). The listener's cognition of sound, which is often spatially separated from its source, creates an "image" of expression. One can be transported to a nostalgic memory, a location from their past, it can trigger the recollection of other sounds, or as Emmerson explains, an "image" containing each of these experiences. In *The Relation of Language to Materials* (1986), Emmerson was not focused on establishing a definition of musical expression, or even to fully uncover what "images" are evoked by music, but rather the relationship between musical composition and the perceived "image" (17).

It is important to understand, not just the way music is potentially able to express meaning, but how the relationship between music and language is reflected in literature. In this regard, four main concepts are commonly used: music as language, language in music, music in language, and language about music (Feld and Fox 1994: 26-27). Music as language details the view that music can be approached using linguistic models of analysis and expression (26). Music in this category is viewed "as an autonomous formal domain, abstractable as hierarchical structure or cognitive process" (26). The structures can be interpreted using linguistic organisational ideas like grammar and syntax, encompassing morphology, the study of how words are formed from

phonemes, and phonology, the study of the sound patterns used in language (26). This approach to understanding music falls in line with the constructivist, structuralist linguistic models that hold the belief that languages are autonomous formal domains. Regarding music as language, gesture should not be overlooked. Language is produced by physical action: the vibration of one's vocal cords, the shape of one's mouth and lips, and often with facial expression and other bodily gestures like hand movement and changes in posture. Music also relies on gesture for its production: acoustic, electric, and digital instruments all rely on gesture to produce sound.

Language in music investigates the "intertwining of language and music in verbal art, song texts, and musical performance" (Feld and Fox 1994: 27). Prosody, the patterning of stresses and intonation used in language, and paralanguage (vocal timbre, the use of dynamic expression, and pacing/tempo) are contained within the realm of music in language (27). Lastly, a focus on language about music investigates the use of language within discourses that surround musical aesthetics (27).

By examining the parameters shared by music and speech such as pitch, rhythm, and timbre, one could hypothesize that the emotive information interpreted in speech is a result of the same "felt" expression of an idea. It is not the linguistic content alone that facilitates our understanding of someone's emotional state during a conversation, but its symbiotic partnership with features of the mode of delivery.

All these categories that consider the relationship between music and language, whether analogous to linguistic models or the language about musical discourse, need to be contextualised and understood from the perspective of being firmly rooted in western ideas of musical form and linguistic thought. Even Langer's interpretation relies on western conceptualisation of sentient experience. Although music and language are universals in the sense that every culture uses them, the theoretical understanding of how they function is far from universal (Feld and Fox 1994: 28). This research does not claim that its findings are universal truths of musical and linguistic meaning, but explored how the results could be applied to creative practice.

2.3 Cut-Up

The cut-up technique was used and popularised by William Burroughs, an American writer who "was widely recognized as one of the most politically trenchant, culturally influential and innovative artists of the twentieth century" (Burroughs 2014: i). Burroughs used forms of the technique in his writing. *The Soft Machine* (1961), *The Ticket That Exploded* (1962), and *Nova Express* (1964), referred to as *The Nova Trilogy* or *The Cut-Up Trilogy*, are three novel-style works that Burroughs wrote using the cut-up technique.

Applying the cut-up technique to audio is something Burroughs was aware of, explaining that Brion Gysin, the "rediscoverer" of the cut-up technique, had "pointed out that the cut-up method could be carried much further on tape recorders" (Odier 1970: 28). Burroughs' use of the method in his critically acclaimed *The Nova Trilogy* (1961-64) means that his name has become intrinsically associated with it, but he was always quick to try and explain that Gysin was the inventor of the idea (Ryan 2022). Burroughs explains that tape recorders offer a variety of ways in which sounds can be manipulated that are impossible to achieve in writing: "effects of simultaneity, echoes, speed-ups, slow-downs, playing three tracks at once, and so forth" (Odier 1970: 29). Burroughs' use of the cut-up technique in his writing influenced the compositional approach and aesthetic of this research.

Burroughs' cut-up method was fluid in the sense that he employed cut-ups to varying degrees in his creative process; using parts of cut-up text as the inspiration for larger, more linear, ideas and discarding the experiments, or by making use of actual sentences comprised of cut-up material (Odier 1970: 29). It was, by his accounts, both a conscious and unconscious endeavour in the sense that he was in control of the material he proceeded to cut-up (although this can be a random act as well) and whichever method he used to rearrange the material, but unconscious of what the result would be (30). The success or failure of a cut-up would be determined upon an examination of the finished experiment and not based on the method chosen to achieve the result (30).

"It is hoped that the extension of cut-up techniques will lead to more precise verbal experiments closing this gap and giving a whole new dimension to writing" (Odier 1970: 27). Burroughs was hopeful that by extending the cut-up method into the realms of music (audio), film and

photography (visual), mediums less limited than the written word, it could help elevate the creative possibilities of writing. Early editing techniques in film, such as cross-cutting, could have possibly been the inspiration for the adoption of similar techniques in other mediums. By manipulating recordings of human speech, the researcher intends to explore the possibilities that the cut-up technique presents to investigate whether the sonic information in cut-up speech sounds aid one's understanding of signified emotive meaning in language.

Cut-ups should not be thought of as a relic of the past, either. In our virtual lives, many of us experience examples of cut-up on a regular basis. Short-form videos on Instagram and Tik-Tok employ a technique known as stitching in which a user can edit video and audio content with existing content on the platform to create their own versions of the original. In fact, Burroughs was fully aware of how powerful a tool cut-up could be in the entertainment industry, stating, "you want the widest possible circulation for your cut-up video tapes. Cut-ups techniques could swamp the mass media with total illusion" (Odier 1970: 181).

2.4 Artistic Research

Artistic research, like any research tradition, is rooted in the desire to ask and answer questions, solve problems, ascertain new knowledge, and add to the wealth of knowledge that precedes it (Nelson 2013: 3). However, the focus of artistic research is that there exists "a special mode of functioning as an artist that goes beyond the natural and intuitive enquiring of the artistic mind and encompasses something of the more systematic methods and explicitly articulated objectives of research" (Crispin 2015: 56).

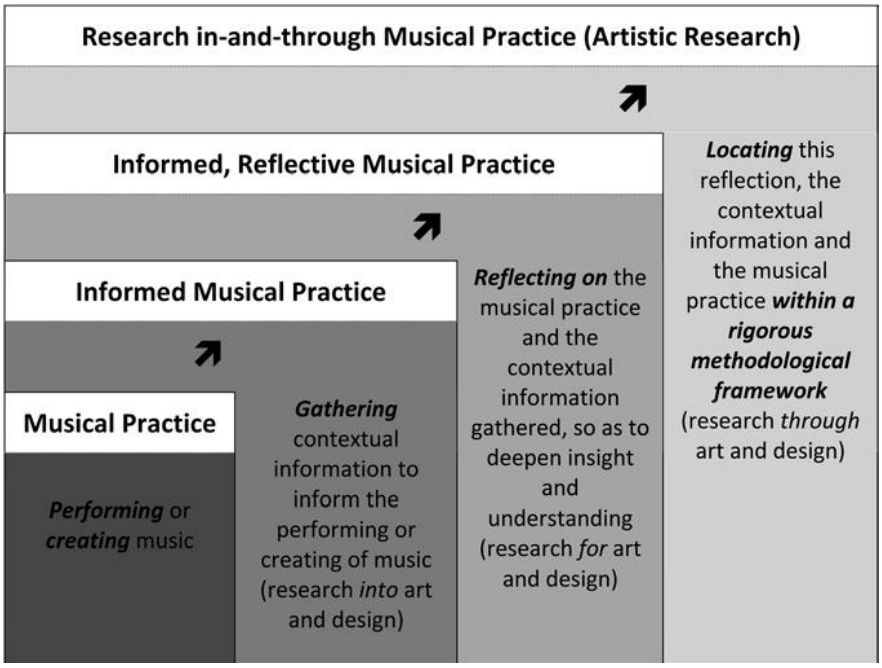
The artist-researcher should try and find a method that allows them to exist between the realms of the purely intuitive and "subjective" exploration and that of "objective" or scientific research (Crispin 2015: 57; Young 2015: 153). One must be able to exist and act as an insider to the creative practice, i.e., the person or participant making the work, and step away from that perspective and occupy the position of observer or critic. Navigation between these two modes is crucial to an informed artistic research methodology (Suoranta, Hannula, Vadén 2014: 16). The researcher's experience of engaging with existing literature and the subsequent analysis of the results derived from the participant responses informed how the creative practice approached the use of cut-up speech sounds. Similarly, the insight gained from the researcher's

creative practice acted as a catalyst for the pursuit of research that could help to conceptualise their experience.

Scholar, Estelle Barrett shares Crispin's view, stating that both explicit and tacit knowledge are required by practitioners of creative arts research (Barrett 2007: 4). Barrett's belief in the validity of creative arts research comes, in part, from its ability to expose new or marginalised "realities" outside of traditional research scopes (4). Artistic research is subjectively motivated, intuitive, and due to the nature of creative practices, multidisciplinary, with insight and inspiration drawn from many disparate sources.

The artist-researcher is positioned to have the ability to pose research questions that would not occur to the scientific researcher due to the process-based nature of artistic enquiry, resulting in the production of knowledge that can produce insight into creative practice (Crispin 2015: 60). However, the question of successful outcomes is always present when discussing artistic practice as research. As Barret writes: "because of the complex experimental, material and social processes through which artistic production occurs and is subsequently taken up, it is not always possible to quantify outcomes of studio production" (2007: 3). To ensure that one can interpret the results of artistic research, it is of vital importance that the research question(s) be focused to clearly convey that which the researcher is trying to investigate (Crispin 2015: 61). The result of the artistic research, apart from a written thesis, is the practice or creative output of the research. This is vital to artistic research as the submitted practice is "evidence of research inquiry" (Nelson 2013: 9). Figure 1 is a visual representation of the process by which musical practice moves from being the core of musical art-making to a process viewed as viable research activity (Crispin 2015: 58). Musical practice is the act of performing or creating a piece of music. Informed musical practice is the act of acquiring information that can be used to inform a musical performance or inform the choices made in the creation of a piece of music. This information could relate to the historical accuracy of a performance (how would a piece have been performed at the time of its creation), or the stylistic/aesthetic choices made by a composer working within a particular genre. Informed, reflective musical practice is the process of reflecting on how the performance or compositional choices, made based on contextual information, influence the music. Lastly, research in-and-through musical practice (artistic research) is the meticulous process of research into the various processes of making music and situating these processes within a methodological framework (57-58). Artistic research should encompass all the categories from which it extends.

Figure 2.1: "From Musical Practice to Artistic Research" (Crispin 2015: 58)



The role of theory is vitally important to the artistic researcher as it can be used to derive the tools to investigate conceptual spaces (the artistic world created by the practitioner) (Young 2015: 163). The process of working within the subjective world of taste, intuition, and feeling, and the objective world of theory and research, facilitates progressive results in the creation of artistic works. The conceptual space, born from a subjective creative practice, can be navigated and conceptualised using theoretical tools (music theory, colour theory, conceptual frameworks such as Dennis Smalley's spectromorphology (Smalley 1997: 107)), providing structure and organisation to a possibly disconnected artistic discourse.

2.5 Popular Music and Composition

This section presents literature pertaining to popular music as the researcher's own musical works have drawn from selected aesthetics and structural characteristics from the popular music discourse. The writings of scholar, Keith Negus, and philosopher, Theodore Gracyk, outline the conceptual frameworks as they analyse narrative and timbre respectively in popular music. This section also presents the researcher's musical influences: the aesthetic and/or conceptual approaches that inspired his musical works.

Negus's paper "Narrative, Interpretation, and the Popular Song" (2012) posits that within the realm of popular music, "even if the inspiration is implicit or unacknowledged, songs are heard alongside and in relation to other songs (by songwriter and listener alike)" (370). Negus implies that part of the meaning behind a song is in the listeners' (including the songwriter's) subjective position. Their interpretation is influenced by their prior interaction with other songs. To analyse meaning in popular music, Negus suggests that one should approach from an inter-subjective and inter-contextual perspective (2012: 368). Inter-subjectivity means that an idea, belief, or understanding of something is shared between two or more people. In the context of music, inter-subjectivity can be thought to represent a shared musical experience or a shared interpretation of musical meaning. Musical inter-contextuality refers to the shared context of a musical event.

The relevance of Negus's approach for the researcher is that it focuses on the way popular music can express narrative meaning. A narrative approach to musical analysis is when one experiences an account of various connected musical events and one's interpretation of the relationship between these musical events is what expresses meaning. Narrative is far more accessible when lyrical content is present in musical works because one is more able to connect semantic events into a logical narrative. However, music without lyrics still presents events that can be interpreted narratively, albeit in a more ambiguous manner. Negus speaks of the "poetic possibilities of ambiguity" (2012: 379), in which popular music's ambiguous narrative expression contributes to its inter-subjective and inter-contextual emotive expression. He uses the example of Randy Newman's "You've Got a Friend in Me" (1995) from the Pixar film *Toy Story* (1995). In his interpretation, Negus explains that in the context of the film the song describes the friendship between a boy and his toy. The context of the film makes the semantic content of the lyrics and musical events very clear, but if a listener were to listen to the song in a separate context, then the ambiguous nature of the events presented in the song are applicable to other relationships (such as personal friendships). The lyrics limit infinite interpretations of the song's meaning, but this illustrates the powerful effect that context and subjectivity have on the listener's experience.

From the perspective of a composer, the understanding that narrative is a crucial part of popular music's expression aided the researcher in conceptualising their musical works. The researcher's approach to each of the eleven tracks in the album created was to develop a metaphysical narrative representing his experience with numerous characters. The tracks were

composed with the idea that each musical event would be interpreted (either consciously or unconsciously) by a listener and experienced as a recollection of an interaction. Negus's description of music's ambiguity and his advocation for an inter-subjective and inter-contextual approach intertwines with the second research question (How does the act of composing with cut-up speech sounds change the emotive meaning for the receiver?). The ambiguous or lack of semantic information in the cut-up speech sounds opens the interpretation of the music to varying inter-subjective and inter-contextual experiences.

Having focused on the conceptual analysis of popular music, what follows is a discussion of the composite elements of popular music and the importance of timbre as a significant contributing factor of the emotional experience of a listener. Gracyk's argument in his paper "Sound and Vision: Colour in Visual Art and Popular Music" (2003) is that "technological developments enhancing the role of timbre in musical arrangements make timbre highly analogous to visual colour in modern painting" (49). This process has allowed timbre to act at a level equal to parameters that are conventionally thought of as defining formal structure (melody, harmony, form, orchestration) in conveying emotive expression in popular music (52). Gracyk compares musical timbre to the use of colour in visual art stating that historically the two elements have been regarded as secondary to the structural elements of composition in both mediums. In relation to rock music, he states that the structural clarity exhibited in many of the works in the genre is in service of the true expressive resources of the music: song writing, production, and tone colour (50).

The researcher's adoption of a popular music aesthetic was influenced by the desire to emphasise timbre as the core expressive parameter in the music. Although both speech and music share the structural parameters of pitch (by extension, melody) and rhythm that can be organised according to linguistic or music syntaxes, the timbre of the cut-up speech sounds presented a rich resource with which to express emotive meaning in a musical context. In experimenting with this idea in an accessible musical context, a popular music approach to structure facilitated an experience that could foreground timbre within established musical structures.

Some of the researcher's musical influences use the ideas expressed above to great musical effect. It must be noted here that any discussion of popular music places one in the realm of subjective interpretation and that Negus states, paraphrasing Chris Kennett (2008), that

"personal listening is all that scholars can legitimately offer" (2012: 377). Considering this, the researcher discusses the artists below only in service of providing context for the aesthetic choices of his own musical works.

Nine Inch Nails' "Hurt" (1994), the final track of the album *The Downward Spiral*, through production and intentional performance, draws attention to Trent Reznor's intimate and strained vocal timbre in delivering a visceral and expressive account. Supported by diatonic melodic phrasing and a stable and unwavering rhythmic foundation (among other structural elements), the timbre of the voice acts as one of the key elements expressing emotive meaning. Aesthetically, Trent Reznor (founder and principal songwriter of Nine Inch Nails) has opted for an unambiguous approach to the structural organisation of his music, in service of timbrally expressive orchestration. This approach results in music that presents a listener with enough established and common musical context that one's attention can be drawn to production, timbre, and narrative of the music.

Amon Tobin, the Brazilian electronic music producer, though far more experimental in his approach to sound design/timbre and arrangement, was significantly influential in the researcher's musical work. Tobin's music served as an aesthetic touchstone throughout the compositional process (specifically *Out From Out Where* (2002)) because his music is novel, exciting, dense, and presented in a way that does not immediately alienate a listener. He "plays" with, stretches, and warps the structural components of popular music in a way that challenges or reframes one's understanding of their function within the music. Tobin has also made use of the cut-up vocal fragments is his music. "Verbal" (feat. MC Decimal R.) (2002) from Tobin's *Out From Out Where* (2002) uses cut-up vocals as the lead element of the piece. His 2011 album *ISAM* is collection of thirteen tracks that were created from sound recordings of his child's voice, manipulated to varying degrees to create timbrally rich musical environments. The similarity of source material that Tobin uses in these pieces influenced how the researcher treated the cut-up speech sounds. The tracks on *ISAM* (2011) often obscure the voice to the extent that it is challenging to perceive the original source sounds, while "Verbal" presents the vocal cut-up in a direct and obvious way so that the listener is immediately aware of the source material. The researcher experimented with the degree to which the cut-up speech sounds could be manipulated to find expressive possibilities within the material but was careful to retain the vocal identity.

Along with cut-up speech sounds, drums were the only other instrument used to create the musical works for this research. As a result, the researcher's approach to the instrument was similarly considered. One influence was Nate Smith's *Pocket Change* (2018), a solo drum album that establishes and develops thematic variation in the context of established pop, funk, and jazz -influenced, groove-based drumming. Smith's variation of repetitive patterns within a groove context was adopted by the researcher in the composition of his drum parts for the album. Subtle changes to the thematic content of groove-based drumming gave the music a stable and predictable foundation that also propelled its development.

JoJo Mayer's work with his electronic band Nerve created an environment in which acoustic drums and electronic instrumentation coexist. Mayer and Jack Quartet's performance of Don Li's "Different Zones" (2021) makes use of drum kit, string quartet, and recorded playback of Mayer's speech. Speech phrases are repetitively played while the drums and string quartet metrically modulate the musical accompaniment to alter the listeners perception of the music. This performance, although not only for drum kit and speech, acted as a reference that demonstrated a specific approach in which the drum kit could be used to accompany recorded speech and how compositional arrangement could be used to alter the perception of repeated phrases.

An aesthetic influence that informed the researcher's music was Radiohead. Aesthetically, Radiohead balance experimentation with established pop tropes. The band has also made use of the cut-up technique in both writing lyrical content and vocal arrangement. "Everything In Its Right Place" from their album *Kid A* (2000) starts with Thom Yorke's vocals cut-up and layered over a synthesized melodic pad. The remainder of the piece contains sung vocals phrases derived from lyric cut-ups while cut-up vocals move to the background and provide rhythmic development and timbral contrast to lead vocals and instrumental accompaniment. The influences described in this section (among others) illuminated potential avenues of creative exploration during the compositional process, some of which were conceptual and bear little resemblance to the original artists, while other influences expose themselves in more obvious and direct references.

2.6 Qualitative Research

This research gathered qualitative data from participants to answer the three primary research questions. Questionnaires were used to facilitate group responses and one-on-one interviews were conducted to gather more nuanced and detailed qualitative data. All the data gathered for this research was organised and analysed using the process of thematic coding, outlined by psychologists Virginia Braun and Victoria Clarke (2013), to locate common themes found in the data and extrapolate associations and trends in the participant responses. Chapter 3 details how thematic coding was used to interpret the responses of participants and explains the quantitative methods used to investigate the responses for instances of significant correlation.

The process of coding is commonly used in analysing qualitative data. It describes the process of relating the data that has been captured to one's initial research questions or objectives by identifying aspects that are relevant (Braun and Clarke 2013: 206). However, it is not always immediately evident to the researcher which aspects of the data are of relevance. Braun and Clarke identify two categories of thematic coding: (1) selective coding and (2) complete coding (206). Selective coding is the process of finding instances that support the researcher's theoretical and analytical goals. The instances that pique interest but are not directly relevant to the initial topic of the research are not coded. In the case of selective coding the researcher is fully aware of the types of responses or data that will be useful for their research aims. Complete coding requires the researcher to code the entire data set, taking note of everything that may be of interest. Once this process is complete, the researcher can be more selective (206). Complete coding ensures that the entire dataset has been explored for any useful insight, allowing the researcher the flexibility to adjust based on the information the data contains.

The codes created during the process of selective and complete coding are named using "a word or brief phrase that captures the essence of why you think a particular bit of data may be useful" (Braun and Clarke 2013: 207). The codes used to group various instances of data together can be data-driven or researcher-driven, depending on the goals of the research being undertaken. Data-driven coding requires the researcher to use semantic codes derived from the responses in the data because they act as a summary of explicit information and mirror the language used by participants in their responses (207). Researcher-driven codes relate to the researcher's own conceptual and theoretical frameworks and function as references to implicit information in the data (207). Researcher-driven coding can be susceptible to confirmation bias

in that the researcher is analysing the data with the intention of categorising the results based on previous theoretical frameworks, which can result in a skewed representation of the responses. Thus, the researcher chose to implement a data-driven coding approach which, while still subjective and vulnerable to the researcher's own bias, designs data categories based on the information in the responses rather than preconceived theoretical perspectives.

There are three common theoretical approaches to using research interviews as a method for gathering qualitative data. The first and second perspectives, described by scholars Sandy Qu and John Dumay in their paper "The Qualitative Research Interview" (2011), neopositivist and romanticism, are representative of more historically established views, whilst the third perspective, localism, looks to disrupt conventional theoretical approaches (Qu and Dumay 2011: 240-241). The neopositivist interview approach treats the accounts of the interviewee, regarded in this approach as a knowledgeable truth teller, as an objective transfer of knowledge, and considers the interview as a tool for obtaining data (241). The potential limitation of this view is that context and the interviewee are removed from the equation. Only the information is considered, and it is considered the truth, spoken by someone who can explain potentially complex or sensitive ideas accurately and logically in an eloquent manner (Koven 2014: 503).

The romanticism interview approach states that "interviews are best understood as speech events" (Koven 2014: 501). They take place in a location, with participants that are dynamic and multifaceted. The romanticism approach regards the interview process as an encounter between interviewer (an empathetic listener) and interviewee (a participant), the accounts of which are viewed as "a pipeline of knowledge mirroring interior and exterior reality leading to in-depth shared understanding" (Qu and Dumay 2011: 241). The romanticism approach is often favoured for ethnographic interviews, while the neopositivist approach is considered more scientific and context-free in nature.

A localist interview method allows a researcher to flexibly approach complex issues from various perspectives (Qu and Dumay 2011: 242) and produces "situational accounts that must be understood in their own social context" (241). Localism resists the idea that interviews are situations removed from social context and place emphasis on interpreting the narrative of an interview as an account of a phenomenon that is situated in an empirical setting (242). The localist approach is considered to occupy the borders between structured, semi-structured, and unstructured interview styles and has the flexibility to function within each of these situations.

3. Methodology

This chapter details the chronological methods chosen during the research process; the subheadings reflect this progressive process. These methods were chosen to answer this project's primary research questions through the codification of participant responses to controlled voice recordings with the aim of creating a taxonomy of vocal sounds which could be used as musical material. A mixed-method approach that draws from artistic research, qualitative data capture, and quantitative statistical analysis was used.

As outlined in the introductory chapter, the researcher aimed to test and answer these questions through the process of musical compositions utilising fragmented, digitally recorded, speech audio. The researcher chose to gather and analyse data collected from participant responses to incorporate the perspective of the listener in artistic practice and ground the conclusions of the research in a mixed-method approach.

What follows is a brief description of each subsection contained within this methodology. It is presented as such to give the reader a broad overview of the methodological process before explaining the application of each method.

Initial Research: Initial research phase and context surrounding linguistic literature.
Creating Examples: The creation of audio examples. Explanation of each audio example and its rationale.
Designing A Questionnaire: Methodology and rationale for the use of questionnaires.
Qualitative Data Gathering: How and why qualitative data was gathered from participants.
Data Capture and Transcription: Data processing for analysis.
Thematic Analysis and Coding: Methodologies used to analyse the qualitative data.
Quantitative Analysis: Analysing the qualitative data using quantitative methods.
Compositional Methodology: Explores this project's compositional process.

3.1 Initial Research

This research sought to understand how the speech sounds used in language communicate emotive meaning. What are the formal elements of one's speech and how do they combine to express emotive meaning?

The field of linguistics was important in the initial research , focusing on phonetics, phonology, and morphology. An understanding of the phonetic sounds of the voice allows for the isolation of speech sounds based on established linguistic definitions. Hypothetically, any conclusion derived from the data gathered by this project could theoretically be applied to other compositions utilising cut-up speech sounds.

The digital recordings used by the researcher in the initial phase of experimentation comprised brief clips of speech describing the researcher's surroundings, emotive readings from books, and expressive phrases attempting to embody feelings including anger, joy, and sadness. The editing process started by identifying phonetic sounds such as vowels and consonants to create audio clips that exhibited one "type" of vocal sound. Some of the experiments contained only the consonant part of a phrase, others were limited to only vowels, and some contained a combination of both.

The aim of creating cut-up speech sounds was to compose clips that utilised the phonetic qualities inherent in the recordings to explore the timbral characteristics of the phonetic sounds in a musical context. The clips containing consonants appeared to have a percussive and somewhat jarring quality/timbre, while the clips containing only vowels appeared to have a more melodic and fluid timbre. The clips containing a combination of vowels and consonants seemed to create the feeling of interplay between the two timbres.

The researcher's initial listening experience was paradoxical in nature. On the one hand, the sounds were clearly vocal utterances that held onto enough sonic information to "hint" at semantic meaning, and on the other hand, the sounds were somewhat jarring due to their fragmented nature and resisted the superimposition of meaning. The juxtaposition was a successful result because, from a compositional viewpoint, the fragmented sounds appeared familiar enough to be understood as vocal sounds but had the potential to draw one deeper into the experience. In the attempt to ascertain any semantic or emotive meaning, the listener would have to "fill in the gaps", thus encouraging interaction.

The subjective nature of the insight gleaned from the initial experimentation is not to be overlooked. It is possible that a different listener, one that was not directly involved in the ideation, recording, and creative act of producing the vocal cut-up audio, may have reacted differently when exposed to the audio. Therefore, to test whether the researcher's interpretation

would be one shared by others, the project set about designing two situations in which qualitative responses to the cut-ups could be gathered through group surveys and one-on-one interviews respectively. The next section of this methodology explains how the cut-up voice recordings, referred to henceforth as "examples", were created.

3.2 Creating Examples

Building on the experimentation described in the previous section, the researcher created a series of cut-up examples to be presented to participants to record and analyse their responses. By analysing responses, the researcher set out to create a musical taxonomy of the cut-up speech sounds that could be used to compose musical works.

The examples were named for the chronological manner in which they would be played for a survey group (Example 1-26). The artistic methodology that was used to create these examples was influenced by William Burroughs and his use of the "cut-up technique" in his series of novels titled *The Nova Trilogy* (1961-1964).

26 examples were created. The number of examples was influenced by the duration of the focus group sessions; the first session was to last one hour, and the participants would need adequate time to complete their responses. Each example was allocated two minutes, totalling 52 minutes of response time and roughly ten minutes for an intermission. The audio was recorded for the sole purpose of this project. The examples were created from recordings of five people of varying gender and ethnicity. The source material was recorded in as similar conditions as possible with the intention to minimise the discrepancy in the recording quality from source-to-source.

Table 3.1: Demographic Description of Sources

Source	Age	Gender	Ethnicity	Home Language
A	24	Female	Coloured	English
B	25	Male	Black	English
C	25	Male	White	English
D	22	Male	Black	Sesotho
E	25	Female	White	English

Sources A, C, and E were recorded in the researcher's home studio with the aid of a vocal booth to minimise ambient noise in the recordings. Sources B and D, for logistical reasons, were unable to record in the same location. However, a brief outlining the recording setup (mic placement and performance notes) and content was supplied. Each source was asked to record a series of phrases in three emotive tones: happy, sad, and neutral. Suggestions were given for phrases, but each source was instructed to supply their own additional phrases if they wished, providing the emotive expression was in-line with the above.

The speech audio was edited by removing one or more parts from the recording, resulting in a fragmented cut-up. Eight parameters defined the characteristics of the examples: pitch; timbre; attack; proximity; rhythm; articulation; contour; and emotion.

Table 3.2: Parameters for Defining Examples

Parameter	Definition	Implementation
1. Pitch	Is the example high-pitched, low-pitched, or neutral?	Cut-ups were edited in such a way as to draw attention to the pitch of the sounds.
2. Timbre	What is the "colour" of the sound?	Digital processing effects were used when necessary to alter the timbre.
3. Attack	Refers to the how the initial transient of the sound wave unfolds over time.	What parts of the words were removed?
4. Proximity	Where is the sound spatially located in relation to the listener?	Source material: proximity to the microphone. Digital processing: delay.
5. Rhythm	Does the example employ a steady or uneven rhythm?	Choosing where cuts were made to create a frantic, steady, or stagnant rhythm.
6. Delivery	In what way was the speech spoken? Whisper, shout, etc.	Using sources that exhibit a variety of articulation.
7. Contour	How does the pitch change over time?	Create examples that exhibit a range of contours. Low-to-

		high, high-to-low, monotone, etc.
8. Emotion	What is the emotional import of the source sound?	Editing and processing in such a way as to either diminish or augment the original emotion of the speaker's delivery.

Although a certain parameter was employed to guide the editing and processing of each example, speech audio inherently exhibits a number, if not all, of these parameters. The researcher is aware that the participants may not be drawn to, or compelled to respond to, the focal parameter of each example.

Table 3.3: Breakdown of Examples[1]

Example	Speaker	Original Phrase	Edited Phrase (spelt phonetically)[2]	Emotion	Parameter
1	A	"Aw, are you sure there is nothing we can do?"	"aw-y-u-uh-e-an-oo"	Sombre	Pitch
2	C	"Yeah, I understand. I just never thought it would happen like this."	"ea-un-an-sne-ou-ha-en-ts"	Sombre	Timbre
3	D	"Who do you think you are?"	"Ho-d-y-th-nk-y-ah"	Angry	Proximity
4	B	"You just don't understand."	"na-ts-nu-ts-uo"	Neutral	Attack
5	A	"I swear to God, if I have to ask you one more time!"	"i-swe-t-go-f-i-ve-t-k-o-mo-t"	Angry	Rhythm
6	C	"Don't you ever say I didn't care. How dare you."	"uo-ea-e-in-d-dy-a-sa-oy-tn"	Angry	Articulation
7	E	"No ways, it's been such a long time."	"n-wa-s-it-s-een-uh-on-I"	Happy	Contour
8	B	"You just don't understand."	"u-on-u-n-er-an"	Angry	Emotion
9	D	"Ke motlotlo ka wena monna." – *I'm proud of you, man.*	"ke-tl-tlo-ka-we-mo"	Happy	Attack
10	B	"You just don't understand."	"dan-e-un-nd-u-nd-st-ju-ne-an-un-dst-a-d"	Happy	Pitch
11	A	"Yay! I'm so happy for you."	"ou-fy-pa-ay-y"	Happy	Articulation
12	E	"Can I tell you a secret?"	"ca-te-se-cr-t"		Proximity

[1] See appendix A1.1 – A1.26 for a comprehensive breakdown of the artistic intention and processing applied to each example.

[2] https://drive.google.com/drive/folders/1Z48UKr37gaqX0TIiaxuhYVOGr5wWEshF?usp=drive_link

13	E	"Yes, I understand. I just never thought it would happen like this."	"sec-klou-is-men-ts-ch-na-nus-at-ts-du-ah-a-si"	Sombre	Emotion
14	C	"Bag of bones that the fastest hand punctuated with fan-fares interspersed even bleaker state of gloom although how could anyone console him."	"ba-bo-st-at-th-fa-tst-a-ctu-ith-f-e-s-per-ev-n-eak-r-sta-of-gl-thou-ow-could-y-c-sole"	Neutral	Rhythm
15	C	"Walking into the night, falling in the pool and yelling for help, but then you realise you are sleeping … f*#k."	"wa-hat-oo-hel-ing-fu"	Neutral	Timbre
16	A	"Oh my gosh, that's so exciting!"	"o-y-o-a-o-e-i-ing"	Happy	Contour
17 (looped)	D	"I'm proud of you, man."	"ou-o-yu-ma"	Happy	Emotion
18	C	"I asked you to do it and now look what's happened! This could all have been avoided!"	"i-ah-d-u-t-now-ook-wh-ts-hp-is-c-d-a-b-n-a-vo"	Angry	Articulation
19	B	"You just don't understand."	"d-st-e-an-ts-ow-ts-oy"	Sad	Rhythm
20	E	"Yes! Go! Woohoo!"	"ye-wo-go-hoo"	Happy	Attack
21 (looped)	D	"Ke masoabi haholo ho utloa seo." *– I'm so sorry to hear that.*	"em-oa-o-tl-se"	Sad	Proximity
22	B	"You just don't understand."	"s-e-an-st-a-an-s-e-an-st-st-a-a-an"	Happy	Timbre
23	C	"Bag of bones that the fastest hand punctuated with fanfares interspersed even bleaker state of gloom although how could anyone console him."	Unintelligible	Neutral	Pitch
24	D	"Ha ke tsebe, na u lekile ho mo letsetsa." *– I don't know, have you tried calling him?*	"ha-ets-a-u-omo-ets"	Neutral	Articulation
25 (looped)	A	"mm mm, no, uh ah, we're not doing that here!"	"m-ah-m-e-w-n-he-oh"	Angry	Emotion
26	E	"I can't believe it."	Unintelligible	Happy	Contour

The phrases were chosen based primarily on whether the emotive delivery was convincing. The examples were assigned a parameter for the editing process based on the inherent attributes of the recordings. The nature of the selection process was subjective, and the researcher understands that the interpretation of the original recordings may not be consistent with other listeners. Examples 9, 21, and 24 use phrases spoken in Sesotho. The reason for doing so was

to determine whether participants would be able to identify cut-up speech sounds of a non-English language.

The recordings were edited using Logic Pro X by selecting clips of the audio and isolating them using the software's cutting tool. The audio clips that remained were edited using Logic's fade tool to ensure that the beginning and end of each small clip was heard without any clicks or pops which would emphasize where the audio was cut.

3.3 Designing a Questionnaire

"Questionnaires offer an objective means of collecting information about people's knowledge, beliefs, attitudes, and behaviours" (Boynton and Greenhalgh 2004: 1312). A questionnaire was created for the participants to supply feedback in the form of qualitative data. It takes the form of a standardised questionnaire, meaning each participant was given the same version of the questionnaire and exposed to the same stimulus (1313). Each example required the participants to answer four questions: (1) How would you describe this example? (2) Was this example happy? (3) Was this example fast? (4) Did you hear a word or a phrase?

(1) How would you describe this example?

This question was designed to give the participants an opportunity to respond in an open-ended manner, allowing for free responses (Boynton and Greenhalgh 2004: 1313). By providing an open-ended question to the participant for each example, the research aims to see how people respond to the audio without prompting. This question occurs first of the four questions because the researcher wished to gather the participant's initial response in as much detail as possible. If placed in a different order the participant may not have had enough time to provide their full insight. This question supplied the bulk of the qualitative data gathered from the participant responses.

(2) Was this example happy?

This was one of two Likert scale questions asked for each example. A Likert scale is "a rating system, designed to measure people's attitudes, opinions, or perceptions" (Jamieson 2023) and is commonly used to gather data for social and educational research (ibid.). The Likert scale offered five responses to choose from: Strongly Disagree; Disagree; Neutral; Agree; Strongly Agree. Both question two and question three are rating scales and therefore produce data that

allows for qualitative statistical analysis (Boynton and Greenhalgh 2004: 1313). The rationale for using a Likert scale was to derive quantitative data that may reveal the extent to which a participant agreed with the assertion that the given example was “happy”. A five-point Likert scale is the most prevalent variation of the scale, as a larger number of points, such as seven, has shown to discourage participants from selecting the extreme positive or negative options, and a four- or six-point scale produces a broad positive or negative response when considering the entire dataset (Jamieson 2023). The prompt, “happy”, asks the listener to focus on the way the audio is delivered and any emotive information present.

(3) Was this example fast?

The Likert scale for this question was used to measure the how the participant’s responded to the question of whether the given example was “fast”. The rationale for using “fast” was to draw the listener’s attention to a parameter of the sound, not the semantic or emotive meaning of the utterances heard. The Likert scale offered five possible responses to the question: Strongly Disagree; Disagree; Neutral; Agree; Strongly Agree.

The Likert scales used for questions two and three supplied the researcher with a means to establish a central tendency for a given example by calculating the mean and mode of the derived data (Jamieson 2023). A Likert scale facilitates this kind of analysis because the five points of the scale are often given ordinal interval values from 1 to 5 (Jamieson 2023). An ordinal level of measurement means that the numbers have directionality: 1 is more negative or positive than 2, and so on (Jamieson 2023). The researcher does not treat the intervals as equal, as doing so implies the assumption that a response of 4 is twice as negative or positive as a response of 2 (Jamieson 2023).

(4) Did you hear a word or phrase?

The last of the questions dealt with whether a participant was able to hear what sounded like a word or a phrase in the example. The question was presented in two parts: (1) Did you hear a word or phrase? (2) If so, what did you hear? If the participant heard a word or phrase, then they would be able to write it out on a separate line. If a word or a phrase was heard, the subsequent analysis could aim to discover the cause of such a phenomenon and draw on the analysis to create similar sonic events in the researcher’s musical works. Part one of question four is a statement-based question in which the participant is asked to respond “yes” or “no” (Boynton and Greenhalgh 2004: 1313), while part two of question four is an open-ended

question that allowed the participant to respond freely, similar to question 1. Hearing the audio in a looped fashion when creating the examples, the researcher was able to interpret words and, occasionally, phrases that were not present in the original recordings. The question was included in the questionnaire to determine whether participants would respond in a similar manner based on brief exposure to the examples.

3.4 Qualitative Data Gathering

To test whether the emotive meaning of a recorded phrase withstood the rearrangement of its syntax (How does the act of composing with cut-up speech sounds change the emotive meaning for the receiver?), the researcher designed two contextual situations in which participant information could be gathered. Once gathered, the qualitative data could be analysed using qualitative and quantitative analysis methods to determine the nature of the participant's experience. Should common themes be interpreted from the participant responses, a taxonomy could potentially be created and used by the researcher in their compositional practice. The degree to which the data could be used to create a musical taxonomy of speech sounds would depend on whether the participant survey and expert interview respondents' responses displayed significant correlation across all 26 examples.

The survey data was gathered from two iterations of group listening. In both sessions the participants were played the 26 audio examples numerous times (the variation in time is discussed below), with breaks in-between each playback for the participants to write their responses. The first iteration of the participant survey comprised non-music expert participants from the public (21 of 45 participants), while the second iteration was conducted with first-year music students (24 of 45 participants), meaning that 53% of the sample set are regarded as having musical training. This bias is discussed in relation to the results (see page 56).

The first listening event was held at The University of the Witwatersrand (Wits) Great Hall in Braamfontein, Johannesburg on the 30th of January 2022. Invitations to the event were sent to contacts via WhatsApp messenger and email. The event took place whilst South Africa was implementing Covid-19 restrictions which limited the number of people that could attend. As a result, the event was not made public, shared, or advertised on social media platforms. Upon accepting the invitation, participants were supplied an information form providing the researcher's aim and how their responses would be used in the research. Care was taken by the

researcher to explain the nature of the audio being presented in a manner that would inform the participant without skewing the data. The Wits Great Hall was a desirable location because of its size. It was large enough to accommodate a group of participants, whilst being able to maintain COVID-19 social distancing regulations. The Great Hall is acoustically treated and houses a built-in PA system that allowed the researcher to play the audio for participants at a sufficient volume while ensuring a high degree of fidelity.

Participants were met at the entrance to the Great Hall and asked to complete a consent form. They were subsequently given a printed copy of the research questionnaire, a black ball-point pen, and a piece of craft board for support whilst completing their questionnaire. Participants were seated in the centre of the hall, with a single seat or more separating them to maintain social distancing. The participants were seated in the centre to ensure the audio was as optimum and consistent as possible from one individual experience to the next. The session lasted one hour, and participants had two minutes per example to answer the given questions. Each example was played five times and looped examples were played four times. Audio playback was at a volume loud enough to be clearly audible by each participant without being uncomfortable. The session had two halves, with a ten-minute intermission after thirteen examples. Once the session had concluded, the participants handed their questionnaires, pens, and boards to the researcher and were given a more in-depth description of how their responses would be used by the research and thanked for their voluntary participation [3].

The second event was held in the Music Room on the 8th floor of the University Corner building at The University of the Witwatersrand, Braamfontein, Johannesburg on the 7th of March 2022. It was required to meet the target threshold of 40 participants, a number that would be an effective sample size for the Central Limit Theorem (CLT) to apply, as the threshold for CLT is a sample of thirty (Ganti 2022). The Central Limit Theorem is "a statistical premise that, given a sufficiently large sample size from a population with a finite level of variance, the mean of all sampled variables from the same population will be approximately equal to the mean of the whole population" (Ganti 2022). The second session group consisted of 24 of the Wits Music Department's first-year class, bringing the total number of participants for the group surveys to 45.

[3] Ethics clearance was obtained through The University of the Witwatersrand.

The Music Room is smaller, with less acoustic treatment than the Great Hall. The researcher acknowledges the inconsistencies this may produce in the data, yet the Music Room offered a space with a permanent PA system, required for the playback of the examples to a large group, space for the participants to listen, and desks to facilitate completion of the questionnaires.

Participants were given a consent form to complete. Upon completing their consent form, they were given a printed copy of the questionnaire. The researcher addressed the group and explained the purpose of the session. The second session lasted half an hour because the entirety of the session was required to be completed during a scheduled first year practical class from 13:15-13:45. Participants in the second group had one-minute to answer the same four questions per example. Each example was played three times and looped examples were played twice. The participants were thanked for their responses and handed in their questionnaires to the researcher and the supervisor. The respondents in both sessions listened to the same 26 audio examples and were given the same questionnaire.

A series of interviews was conducted to gather data from music industry professionals. The distinction between music industry professionals and the first-year music students relates to their years of practical experience working within the music industry and were chosen based on their expertise within specific fields of the industry such as recording engineering, composition, performance, music journalism, and research. It is of interest to this research to understand how a music industry professional would respond to the same set of audio examples, given their experience with the nature of music production. This project includes responses from eight professionals which were gathered through eight 45-minute, one-on-one interviews. Interviews were the chosen method of data gathering to facilitate a more conversational discussion.

The interviews followed a semi-structured format, defined as a qualitative method that uses "prepared questioning guided by identified themes in a consistent and systematic manner interposed with probes designed to elicit more elaborate responses" (Qu and Dumay 2011: 246). The semi-structured interview is the most common qualitative interview method because it is considered flexible (it can be tailored to fit a range of research criteria, but importantly, it succeeds in enabling interviewees to respond in an open manner) (246). As discussed in Chapter 1, there are three common theoretical approaches to the use of research interviews as a method for gathering qualitative data. Neopositivist and romanticism, are representative of

more historically established views, whilst the third perspective, localism, looks to disrupt conventional theoretical approaches (240-241). The researcher chose to adopt a localist approach for the interviews and, specifically, how the responses were interpreted. Responses interpreted from a localist perspective take into consideration the context of the interview and the interviewer-interviewee relationship and are understood as highly variable interactions that produce data that is reflective of a moment and place in time and space. The same responses may not have been gathered if the context surrounding the process was different.

The interviewer (researcher) used the research questionnaire as the guide for the interviews. The participants became familiar with the questions as the interviews unfolded and were not discouraged from answering them out of order or without prompting. This was to maintain the immediacy of their impressions and a conversational atmosphere. On occasion, when the researcher felt that a participant's response could be expanded and sufficient time remained prior to the next example, probing questions were asked to direct the conversation to derive more detailed responses.

A total of six interviews were conducted over two consecutive days, the 28th and 29th of March 2022, at The South African Broadcasting Corporation's (SABC) anechoic chamber in Auckland Park, Johannesburg, and two at a private residence in Melville, Johannesburg on the 25th of April 2022. The interviews were conducted by the researcher and structured so that each audio example was played back in the same manner as the group surveys. Due to the time constraints of each interview session, the examples were played three times and looped examples were played twice. During the playback breaks, the interviewer asked the same four questions (see page 31-33) to the expert respondents. The interviews were audio recorded.

Both the survey sessions and the interviews were time sensitive with participants limited to a set amount of time to respond to each example. The one-on-one interviews each lasted 45 minutes. The interview sessions lasted for an hour each, with the first fifteen minutes dedicated to welcoming the participant and explaining the process, answering questions, and establishing comfort for the interviewee. The hour session was chosen because it is the researcher's belief that any longer might negatively impact the participant's workday.

3.5 Data Capturing

The survey data, once collected, was transcribed verbatim in Microsoft Word, and tabulated digitally in Microsoft Excel. Sheets were created in an Excel workbook that corresponded to the audio examples played during the group listening sessions. This allowed for a streamlined workflow and a more advantageous, digital form of cataloguing the data from the individual questionnaires. To retain anonymity and prevent bias during the analysis, each participant is identified by a numerical value from 1 to 45. The interview data was also captured in Word and Excel as above.

The researcher isolated nouns, adjectives, and verbs from the participant responses to generate preliminary data-driven categories that would become the themes used to code the qualitative data. Similar words were grouped together to create eight categories: sonic descriptions; anthropological descriptions; imagery; music; speech, technology; emotion; and other, for both the group surveys and the expert interviews. The categories are the same for the surveys and interviews as the responses were of a similar nature in both cases. Isolating the nouns, adjectives, and verbs from the original responses could potentially lead to a misinterpretation of the contextual meaning of the words, so careful attention was paid in situations where a word could denote several interpretations.

3.6 Transcription

The interviewee responses were recorded using Logic Pro X, a large diaphragm condenser microphone, and a portable audio interface. The audio from the Logic session was bounced (converted) into a ".wav" audio file and subsequently transcribed.

According to Braun and Clarke in their book *Successful Qualitative Research* (2013), it is of vital importance when transcribing an interview that the language content and speaker are clearly recorded (163). Braun and Clarke also recommend that a good orthographic transcript includes "all verbal utterances from all speakers, both actual words and non-semantic sounds" (163). This research does not "clean up" or edit the data from the interviews, as doing so would make the respondents sound more like they were writing than speaking in a natural manner.

A transcription is not the actual interview that was conducted, rather it is a representation of the event (Braun and Clarke, 2013: 162). The transcription is two-steps removed from the interview as it presents the information gathered by an audio recording of the primary event.

In each step of the process, the recording and the subsequent transcription, data is lost or changed, and it is therefore crucial to remember that the data derived from a transcription should not be considered raw data. Braun and Clarke suggest thinking of a transcription as partially representative data, "already prepared and slightly altered from its original stage" (162).

There was undoubtably nuance and subtlety lost in the transcription and analysis of the interviews. However, all eight transcripts were undertaken in an orthographic style to include as much of interview process as possible. This researcher's aim is to enlighten the semantic meaning of the responses supplied by the interviewees, not to detail the phonetic or paralinguistic features of their responses. A transcript that aims to do so is called a Jeffersonian transcript and is often used in the fields of discursive psychology and conversation analysis (Braun and Clarke 2013: 162).

3.7 Thematic Coding and Analysis

Thematic coding was used to analyse the survey data and interview transcripts. This is as a flexible qualitative analytic method suitable for almost any research question (Braun and Clarke 2013: 177-178). Thematic analysis or TA stands out from other qualitative methods of analysis because it only prescribes a method for analysing data and not the means of collection, the theoretical positions, epistemological or ontological frameworks (178). This allows for a more tailored approach to analysing qualitative data. Braun and Clarke outline two ways themes can be generated from the data: in a bottom-up or top-down fashion (178). Bottom-up themes are derived directly from the data. Top-down themes are based in existing theoretical frameworks the researcher may want to explore in the data, or in the analysis (178). This research employs a bottom-up approach to generating themes from the survey and interview transcripts.

NVivo, created by QSR International, was used to thematically analyse the data. NVivo is known as a computer-assisted qualitative data analysis software (CAQDAS) program. "CAQDAS programs assist qualitative researchers to collect, organize, analyse, visualize, and report their data" (Dhakal 2022: 270). The NVivo program is designed for researchers gathering and analysing qualitative data and is suitable for mixed-method research methodologies (270). Coding in NVivo specifically refers to the process of labelling and

categorising parts of data within the larger dataset (270). NVivo allows a researcher to create the above-mentioned categories of data called codes in a hierarchical system in which multiple codes, each pertaining to a broader code, can be group together (270-271).

Once the coding process had been carried out across the dataset, the researcher was able to generate numerical information about a particular theme, compare the themes across larger organisational groups known as cases, and establish trends in one's data that would otherwise be extremely time consuming. Annotated examples of how NVivo's features were used will be included in Chapter 4.

In a similar way to how transcripts are a twice-removed representation of the original interview, encoding software generates another representation that is further removed from the original data. The researcher has approached NVivo as a tool to (1) understand any thematic connections present in the responses of the participants, knowledge that informed the researcher's artistic practice, and (2) aid in deriving numerical representations of which themes were most present in an example, a process required to analyse the data using quantitative methods.

3.8 Quantitative Analysis

This research investigated the potential connections between participant responses through quantitative analysis methods as a secondary focus. It investigates the correlation between variables in the data which allowed the comparison of two or more variables and an understanding of the degree to which they are related. This allowed the researcher to discover any connections between participant responses and subsequently facilitated the creation of a taxonomy of cut-up speech sounds based on the results. Correlation can be extremely helpful in analysing large amounts of quantitative data for patterns and behaviour, but it does not detail which variables are causing the correlation and what the effect is as a result (Kafle 2019: 127). Correlation analysis is commonly used in the fields of medicine, social sciences, and education (126).

NVivo's strength lies in its ability to generate representations of the themes present in given data. Data can be interpreted through graphs; mind maps; spider diagrams; and as tabulated

numerical data. However, this research also endeavoured to establish if there is a statistical correlation between participant responses. NVivo was used to generate numerical representations of the references within a selected code. The codes containing the highest numerical values across the entire dataset (surveys and interview transcripts) were gathered for each audio example. Seven codes were collected: joy; fear; anger; production (the participants' responses made mention of the editing/production process of the audio); rhythm; tone colour; and technology. As a result, each example possessed seven numerical data points that correspond to the themes coded in NVivo.

Statistical Package for Social Science (SPSS) is a program that is specifically designed to carry out correlation analysis on a dataset. SPSS allows a researcher to find significant statistical correlation between two or more variables and supplies tabulated representations of any correlation that is present based on the selected correlation method. The types of correlation methods that are available to a researcher in SPSS are: Karl Pearson's correlation coefficient (Blythe 1994: 393), Spearman's rank correlation (1904: 79), and Kendall Tau (1938: 81-93). Both Spearman's rank correlation and the Kendal Tau method are used to find correlation in ranked data, or data that has been analysed and given a "place" relative to the other data entries. The researcher made use of Spearman's rank correlation as the intention was to rank data according to the number references to the thematic code and explore the degree to which participants responded similarly to the examples.

Here it is important to define what is meant by association and correlation. Association happens when an individual is presented with two variables and their respective values adjust together. "Association is described in terms of patterns present in the scatter plot" (Kader and Franklin 2008: 293). A scatter plot is a tool used to visualise statistical data and contains points for each data entry. The patterns on the scatter plot express the directionality, strength, and form of the associations between variables. Associations are considered positive or negative depending on the behaviour of the values of each variable and if two variables each display above-average values occurring together, the result is a positive association. Whereas, if one variable displays above-average values whilst another variable displays conversely below-average values, the two are said to have a negative association (293). Understanding how associations organise and interpret quantitative data is beneficial as predictions can be made about the tendencies of one variable's values in relation to others in the dataset (293).

Correlation analysis is the statistical process by which the degree of relationship between variables is exposed (Kafle 2019: 127). “Quantities that measure the strength of association are called correlation coefficients” (Kader and Franklin 2008: 293). The correlation coefficient determined the degree of relationship between examples by comparing the values in each of the seven themes for example A with the same values for example B. If the values behaved in a similar manner across the seven data points, the correlation between the examples analysed was deemed statistically significant.

The correlation coefficient between two variables is measured between -1 and 1. Similar to how positive and negative associations are defined, correlation is positive if the trend of both variables’ values is positive and negative if the values behave conversely (Kader and Franklin, 2008: 297). A value between -1 and 0 is deemed a negative correlation, whilst a value between 0 and 1 is deemed a positive correlation. The table below provides the values needed for significant correlation:

Table 3.4: Interpretation of Correlation Coefficient Values [4]

Result	Interpretation
Correlation Coefficient = 1	Perfect positive correlation
Correlation Coefficient = -1	Perfect negative correlation
Correlation Coefficient = 0	The variables are uncorrelated
0 < Correlation Coefficient > 0.4	Low positive correlation
-0.4 < Correlation Coefficient > 0	Low negative correlation
0.4 < Correlation Coefficient > 0.7	Moderate positive correlation
-0.7 < Correlation Coefficient > -0.4	Moderate negative correlation
0.7 < Correlation Coefficient > 1	High positive correlation
-1 < Correlation Coefficient > -0.7	High negative correlation

The statistical correlations that are generated by SPSS derived useful insight into how participants responded to the examples, but do not pinpoint the specific reasons for such a correlation to exist. The method that this research employs, deriving the most prevalent themes from the data and analysing these for correlation, means that the nuance in the responses was not fully considered when searching for statistical correlation. This process resulted in

[4] *Table adapted from Dhakal 2022: 128.*

numerous accounts of significant correlation between examples, but when further analysed based on their inherent parameters, it was difficult to ascertain the precise reasons for correlation.

3.9 Compositional Methodology

Instrumentation/ Source Material

This research made use of acoustic drum kit, a sample pad which was used to interact with Ableton [5] software instruments, and recorded voice. The drum kit and the cut-up speech sounds have contrasting timbres which lend depth to the music. This both aided and challenged the compositional process. The ability to contrast the melodic nature of the vocals with percussive sounds was a valuable tool when developing and providing variation to established themes but was problematic when the two timbres needed to inhabit the same space cohesively. Where necessary in the pieces, digital effects were added to the drum kit audio to make the sounds congruous.

The cut-up speech sounds used in the musical works were taken from the same phrases used to create the examples for the data-gathering process. The use of the recorded voices from the earlier stages of the research further linked the research and artistic output. The information gathered from the participant responses guided the choices made in terms of selection and application of samples. The sources used exhibit a variety of timbres as the five subjects each possess unique voices and the recorded phrases were delivered with varying emotional expression.

Various parameters of the cut-up speech sounds were considered when choosing which sounds to use in certain works. Questions including: "Is the sample percussive, melodic, or timbrally interesting? and “Which fragments pair/group well with others?” guided the choice. Establishing these labels and categorising the fragments was predominantly straightforward and intuitive. The cut-up speech sounds either contained mostly percussive information produced when pronouncing consonant vocal stops, or contained vowel sounds more melodic in nature, like the timbral qualities of a brass or woodwind instrument.

[5] Ableton Live is a Digital Audio Workstation (DAW)

Stops are “when speakers totally block the vocal tract so that air pressure builds up behind the closure” (Simon and Kunkeyani 2014: 94). Vocal stops comprise two categories; plosives when air is released quickly, and affricates when air is released slowly (94-95). The samples that contain stops are similar in timbre to the sounds produced when striking the elements of a drum kit, which suggested a compositional use for their application as a means of expanding the sonic options of the drum kit. The vowel sounds were often used as the building blocks of the melodic and harmonic content of the pieces. The vowel sounds are generally longer (even if only by a few fractions of a second) which supplies more raw information to manipulate but most importantly, the sounds that contain significant vowel sounds appeared analogous to the tone of a sung vocal – the vowels were natural melodic carriers.

The approach to the drum kit in terms of vocabulary was often influenced by the cut-up speech sounds chosen for a particular piece and draws heavily from funk and electronic dance music drumming styles. However, an alternate scenario was also explored to generate thematic content. The free improvisation pieces that began the process (see tracks “Confusion Kills” and “Inner Piece”), feature drums free of any prolonged regular pulse that respond to the sample phrasing, rhythm, and timbre. In these pieces the drums are largely monophonic and strive for a more melodic accompaniment. As the method shifted to create pieces that were more influenced by popular music, a decision that aligned more with the researcher’s skillset and aesthetic taste, the drums took more of a supporting and structuring role like their stereotypical function in this idiom. There was still considerable improvisation, but the result was much more groove-based, repetitive, and the drums take less of a co-lead instrument role. Timbrally, the tone of the voice influenced how the drums should be treated regarding tuning and effects, so the instrumentation could feel connected in both space and function to deliver the desired aesthetic.

Constructing Sampler Instruments

Several iterations of sampler instruments were created for the musical works, each with the intention of solving a creative problem posed by its predecessor. The first iteration featured a single sample for each of eight available trigger-pads. This iteration, upon integration with the acoustic drum kit, familiarised the user with the mechanics of playing both instruments (drum kit and sampler) in real-time. Mechanically, the combination of the two instruments resulted in a more limited interaction with the drums than was desired. However, the intention of the

hybrid setup was an exploration how the two instruments could work in concert, and the process began to expose the types of phrases that would be possible within these self-imposed constraints. Compositionally, the ability to play more samples was needed to generate new ideas and sustain a feeling of development throughout.

Experimentation led to the creation of a randomising Musical Instrument Digital Interface (MIDI) instrument in Ableton Live that would select from a pre-programmed collection of samples (normally 5-7) being triggered by MIDI note information. The intention for using a randomising MIDI instrument was to help generate permutations of the samples that may not have otherwise resulted due to playing tendencies or limitations. Each pad on the sampler would have several timbrally similar cut-up speech sounds that would be randomly played on contact. The random function of the sampler instrument produced sketches that were very dense and chaotic in vocal content but lacked the feeling, structure, and intention needed for meaning. The application of the random sampler instrument was best suited for brief sections of pieces that required a burst of energy or as a way of disrupting an established and thoroughly stated theme.

Sketches were also conceived using this sampler instrument without the random function enabled, which resulted in the ability to play the same melodic or percussive phrase/configuration and vary the rhythmic structure using a single pad. The effect was more akin to how a sequencer cycles through its programmed material. However, this allowed one to predict what sound would play and when, which set up expected sample triggering outcomes during the improvisations.

Exercising control over the expressive parameters of the samples became necessary because each time the sample was triggered, it would play in the exact same way irrespective of the interaction with the controller. In the context of a live performance of the music, the researcher wanted the samples to behave in an analogous manner to that of an acoustic instrument. The velocity MIDI parameters were mapped (in Ableton) and programmed to control a frequency filter which was placed on the relevant samples, thus enabling timbral variation and creating phrases that felt as if they were more dynamic and expressive.

At this stage, much of the sample content remained in brief fragments that lasted for no longer than a second. The desire to create longer and richer samples influenced the decision to explore

the effects placed on the samples within the software instrument. The samples were manipulated to produce the effects of stretching and warping. The sound created lengthened the duration of the sample which satisfied its most basic function, but also added a layer of rhythmic information that aided in moving the feeling of musical time away from a rigid, purely quantised space. The samples were interesting as stand-alone sonic artifacts and gave the feeling of depth to the music. The risk of overly manipulating the samples using time-stretching and warping is that the vocal samples begin to sound like something else entirely. The intention was to retain the sonic properties of speech so that their application in a musical context could be explored, therefore, as interesting as the sounds were that resulted from extreme warping, the use of time-stretching and warping was attenuated.

Generating Ideas

Compositional development began with a lengthy period of free improvisation using a hybrid setup of the acoustic drums and the multitrack sampler interfacing with Ableton Live. Initially, improvisation was used to explore the possibilities of the hybrid instrumental setup but developed into a method for generating compositional material that could be refined into a complete piece. It was used to start the process, create organic form on blank space, and as a tool to add material to existing structures. This period was crucial in highlighting the limitations of the setup, revealing one's tendencies as a player and improvisor, and resulted in pieces that could be judged compositionally. While addressing questions such as: "Is this functioning properly?", "What more could be done?", and "Is this too much?", the intention of the improvised pieces was also to develop a symbiotic relationship between the drum kit and the samples – to explore how the sounds could work together in a cohesive expressive manner. This process extended the organology of the drum kit creating a hybrid, or augmented, instrument that produced both acoustic and sampled sounds. This kind of augmented drum kit configuration is common in popular performance, with drummers often being required to trigger accompanying tracks for playback or making use of sample pads to trigger electronic drum samples.

As the experimentation continued, the process generated ideas using programming (piece by piece working within Ableton) to expand the sonic capabilities of the pieces, unencumbered by improvisational limitations and mechanical abilities. This meant the researcher was able to

exercise nuanced control over the morphology and expression of the cut-up speech sounds to service the development and progression of the music.

Much of the above material served as the foundation of the pieces created, but the lead themes and melodies that came out of this process lacked the structure and thought to make them sufficiently engaging to sustain a whole piece. Thus, melodies were created using the recorded phrases' original audio (used for the examples). The cut-ups were rhythmically and melodically analysed so that the contour and the cadence of the phrase could behave as the melodic lead and focus of the pieces.

Arrangement

Initially the researcher endeavoured to compose long-form works influenced by improvisation and electroacoustic music composition. While some of the material created via improvisation was used for thematic content in the final works, the process made the researcher realise that he wanted to focus on more defined structural design for most of the musical works on the album. Timbre became the driving parameter with which to approach the arrangement of the cut-up speech sounds and clarity of form, melodic structure and harmonic approach were adopted to support the emotive expression of the music. The arrangement of ideas and material sourced during the improvised period was structured according to the innate form that resulted from the improvisations or developed from programmed phrases, predetermined based on established musical forms influenced by popular music. Natural pauses and breaks suggested sectional changes, while increased intensity and density suggested climax and tension that could be countered with juxtaposing material or elaborated and developed to add to the climax or tension.

4. Presentation and Discussion of Results

Here the researcher considers the analysed results of the participant survey respondents' and expert respondents' responses with the aim to provide a comprehensive understanding of how these results speak to the primary research questions of this dissertation (see below). The results are analysed so that insight gleaned from the process and results of the data gathering are made clear.

Primary Research Questions:

1. Do speech and music share parameters that express emotive meaning in an analogous manner?
2. How does the act of composing with cut-up speech sounds change the emotive meaning for the receiver?
3. Can a composer extract and harness the speech sounds which allow one to signify emotive meaning through spoken language?

Before presenting the results, it is important to restate how the data had been organised as these categories govern how the data is presented. On the back of the transcriptions, categories were created that would inform the themes used to code the responses in NVivo. The researcher organised the nouns, verbs, and adjectives used by participants into groups based on their denotative and connotative meanings. This process, although thorough and methodical, was subjective and the researcher understands that many of the words used have a variety of meanings based on their context. Therefore, care was taken to derive the meaning of the words from the context of the participants' responses. The categories highlighted broader themes that were present across the data that could be used to organise the participants' responses for further analysis.

4.1 NVivo

NVivo allows a user to upload files in the form of documents (MS Word, PDF), pictures, audio, and video. Once uploaded, these files can be coded by selecting relevant parts of the file and assigning them to codes. The codes store each reference and make it possible for a user to see every instance of a particular code across the entire dataset. Cases can also be used to code parts of the files that are consistent throughout. The researcher used cases to organise the

responses pertaining to each example (titled Example 1-26) and participant information (titled P 1-45). Attributes were given to these cases so that more nuanced information could be collected. Each participant survey respondent was given attributes based on their age, gender, and home language (collected from their questionnaires). Each example was given attributes based on the gender of the source (male or female), the source of the voice (A-D), whether it was looped (looped or not looped), and the emotional delivery of the original phrase (happy, sad, angry, or neutral).

The themes used to code the participant survey respondents' responses are as follows:

- **Anthropological Descriptions**: includes codes "Age", "Gender", "Languages", "Locations", "Occupation", "People".
- **Emotions**: includes codes "Anger", "Anticipation", "Fear", "Joy", "Loathing", "Sadness", "Surprise".
- **Imagery**: includes codes "Nature", "Place", "Position", "Size", "Spiritual".
- **Music**: includes codes "Form", "Genre", "Melody", "Production", "Rhythm", "Tone Colour".
- **Outliers**
- **Situations***
- **Sonic Descriptions**: includes codes "Cut-Up", "Movement", "Speed", "Texture", "Tone", "Unclear", "Volume".
- **Speech**: includes codes "Mouth Sounds", "Tone of Voice".
- **Technology***

The themes used to code the expert respondents' responses are as follows:

- **Anthropological Descriptions**: includes codes "Age", "Gender", "Languages", "Locations", "Occupation", "People".
- **Emotions**: includes codes "Anger", "Anticipation", "Disgust", "Fear", "Joy", "Love", "Neutral", "Sadness", "Surprise".
- **Imagery**: includes codes "Nature", "Place", "Position", "Quantity", "Size".
- **Music**: includes codes "Engineering", "Form", "Genre or Style", "Melody", "Rhythm", "Tone Colour".

- **Other**
- **Situations***
- **Sonic Descriptions**: includes codes “Cut-Up”, “Movement”, “Speed”, “Texture”, “Tone”, “Unclear”.
- **Speech**: includes codes “Mouth Sounds”, “Tone of Voice”.
- **Technology***

The codes marked with (*) do not contain sub-codes. This means that the references to these themes did not, in the view of the researcher, conform to more nuanced themes that fall under the main code, as was the case with the other main codes.

The above codes were used to organise the responses to the question “How would you describe this example?”.

The responses to the Likert and phrase questions were coded as follows:

- **Happy Likert**: includes codes “Strongly Agree”, “Agree”, “Neutral”, “Disagree”, “Strongly Disagree”.
- **Fast Likert**: includes codes “Strongly Agree”, “Agree”, “Neutral”, “Disagree”, “Strongly Disagree”.
- **Phrase**: includes codes “Yes”, “No”, “Did not answer”.

The discussion begins with the results of the NVivo thematic coding and elaborates on the visual representations of the qualitative data. The responses of the participants and experts are explained and compared to show similarities and differences in their experiences. Following this discussion are the results of the correlation analysis carried out using SPSS. The analysis of quantitative data in SPSS was used to find potential instances of significant correlation between examples. These results show which examples displayed significant correlation across the sample set to understand which characteristics of the audio potentially influenced the participants to respond in an analogous manner.

Lastly, it is important to highlight that the results and discussion are the proposed outcomes of thematic coding, and that the process is highly subjective in nature: the generation of themes

is the result of the researcher's own interpretation of the responses. The usefulness of the results is directly connected to what artistic and compositional approaches they suggested.

4.2 Participant Survey Results

Table 4.1 presents a representation of the ages of the participants involved in the survey process. From this, one can see that most of the participants fall into the range of 18 (end of high school) to 29. As other ages are not adequately represented, the researcher cannot speculate on the relevance of the results for people over 30 years-of-age. 24 of the 45 participants for the group surveys were part of a 1st year Wits University music class, the common age of which are students recently finished with high school (approximately 18) to students in their 20s.

Table 4. 1: Participant Survey Respondents' Age Range

Participant age range	**Number of participants per age range**
20-29	23
18-19	13
30-39	4
50-59	3
17	1
Unassigned	1

Table 4.2 shows the gender of the participants and displays 58% male (26/45), 40% female (18/45), and 2% unassigned (1/45).

Table 4. 2: Participant Survey Respondents' Gender

Gender of participants	**Number of participants per gender**
Male	26
Female	18
Unassigned	1

Lastly, Table 4.3 presents the reader with the home languages of the participants to better understand the potential language and cultural differences with which people experienced the examples. 73% of the participants listed their home language as English, therefore, as was stated regarding participant age, the inadequate representation of participants with other home languages means the relevance of the data for home language groups other than English cannot be speculated upon.

Table 4. 3: Participants Survey Respondents' Home Language

Home Language of participants	**Number of participants per home language**
Afrikaans	2
English	33
IsiZulu	5
Sepedi	1
Setswana	1
Sotho	1
Tswana	1
Unassigned	1

4.3 Coding

Table 4.4 displays the number of references to each code across all 26 examples and were coded from the participants' response to the first question of the questionnaire: "How would you describe this example?". The most common codes referenced were "Anthropological Descriptions", "Emotion", "Music", "Sonic Descriptions", and "Speech". The results displayed in Table 8 show the range of themes referenced for each example and allowed the researcher to determine the most prevalent themes for each example. In terms of the researcher's creative practice, the coded response to each example supplied the researcher with

insight into what associations a listener may have when presented with similar phrases in the musical works.

Table 4. 4: Participant Survey Respondents and All Codes

	A : Anthropological Descriptions	B : Did not answer Q.1	C : Emotions	D : Imagery	E : Music	F : Outliers	G : Situations	H : Sonic Descriptions	I : Speech	J : Technology
1 : Ex 1	7	2	15	17	17	1	2	25	7	2
2 : Ex 2	8	1	14	4	26	2	0	15	5	14
3 : Ex 3	11	2	13	2	6	1	2	21	23	0
4 : Ex 4	7	1	22	4	13	0	2	15	14	3
5 : Ex 5	9	0	22	5	13	3	1	39	5	5
6 : Ex 6	11	0	36	1	6	3	3	6	24	1
7 : Ex 7	17	4	23	0	3	2	0	25	12	4
8 : Ex 8	4	1	24	13	39	1	2	12	13	2
9 : Ex 9	10	2	9	2	15	1	1	34	8	14
10 : Ex 10	35	2	10	0	8	1	3	5	25	1
11 : Ex 11	14	1	26	2	26	0	2	16	6	4
12 : Ex 12	5	0	18	7	38	2	2	17	5	3
13 : Ex 13	18	1	14	0	6	0	3	13	21	11
14 : Ex 14	4	0	17	1	20	1	1	28	10	2
15 : Ex 15	5	1	27	15	3	2	4	5	23	2
16 : Ex 16	12	0	14	6	50	0	1	9	4	2
17 : Ex 17	10	2	15	4	28	0	1	18	4	4
18 : Ex 18	7	3	16	1	3	0	2	29	10	17
19 : Ex 19	3	0	24	7	16	2	2	19	18	12
20 : Ex 20	13	1	42	3	7	2	1	5	7	0
21 : Ex 21	5	5	13	3	36	0	1	20	5	4
22 : Ex 22	12	5	5	6	11	1	1	32	8	6
23 : Ex 23	2	2	18	5	24	0	3	28	5	2
24 : Ex 24	14	1	9	0	4	0	3	16	26	9
25 : Ex 25	14	2	10	2	42	0	0	11	12	1
26 : Ex 26	3	2	19	12	28	1	4	20	3	8

The categories: "Anthropological Descriptions" (how young or old the voice sounded), "Gender" (whether the voice was male or female), "Language" (the language spoken), "Location" (global voice location or an association with a particular place), "Occupation" (the work that is associated with a particular kind of voice), and "People" (nationality) all show that the participants were seemingly trying to contextualise the cut-up speech sounds based on the sonic information available.

The large-scale reference to "Emotion" demonstrates that participants were drawn to how the cut-up speech made them feel, or the emotion they perceived the voice was feeling. This speaks to Langer's proposal that music and one's experience of sentience share a "similarity of logical form" (Langer 1942: 228). The participants were not asked to describe the examples using emotive language, merely for their description of the example. This speaks to the first research question: Do speech and music share parameters that express emotive meaning in an analogous manner? The frequency and quantity with which participants referred to emotion across the

examples shows that when presented with the rhythm, timbre, and pitch of speech (parameters shared by music), an emotive association or sentient experience was triggered.

Considering the frequency with which the theme “Music” appears, the fact that 53% (23/45) of the participants were university music students means that they would possess musical knowledge and musical vocabulary that non-musicians do not. As a result, the relevance of the participants’ responses to music is not representative of how non-musicians would respond.

“Sonic Descriptions” relates to the second research question of this dissertation: How does the act of composing with cut-up speech sounds change the emotive meaning for the receiver? References to: “Sonic Descriptions”, “Cut-Up”, “Movement”, “Speed”, “Texture”, “Tone”, “Unclear”, and “Volume”, demonstrate to the researcher that when semantic and emotive information are not immediately discernible, one’s response is of a descriptive nature. The focus shifts away from understanding what is being experienced at the level of signification, to the spectromorphology (Smalley 1997: 107) of the sound’s physical characteristics. These descriptors can signify meaning beyond the mere characteristics of the sound depending on a participant’s association with speech. However, speculation on this would be vague and presumptuous as the responses of each participant would be varied and deeply rooted in their personal experiences.

Table 4.5 presents references made by participant survey respondents to the “Mouth Sounds” or “Tone of Voice” of the examples. Most references to speech were comments about the “Tone of Voice” of the example, with only examples 8, 15, and 19 displaying references that favoured “Mouth Sounds” . The readiness of the participants to respond to the tone of the voices shows that, much like the references to “Sonic Descriptions”, participants can describe characteristics of the sounds. Unlike references made to “Sonic Descriptions”, references made to “Tone of Voice” represent information about the type of expression that remains perceivable in the voice.

Table 4. 5: Participant Survey Respondents and "Speech"

	A : Mouth Sounds	B : Tone of Voice
1 : Ex 1	3	4
2 : Ex 2	1	4
3 : Ex 3	0	23
4 : Ex 4	0	14
5 : Ex 5	0	5
6 : Ex 6	1	23
7 : Ex 7	0	12
8 : Ex 8	7	6
9 : Ex 9	0	8
10 : Ex 10	0	25
11 : Ex 11	1	5
12 : Ex 12	0	5
13 : Ex 13	1	20
14 : Ex 14	0	10
15 : Ex 15	12	13
16 : Ex 16	0	4
17 : Ex 17	1	3
18 : Ex 18	0	10
19 : Ex 19	10	8
20 : Ex 20	0	7
21 : Ex 21	0	5
22 : Ex 22	1	7
23 : Ex 23	1	4
24 : Ex 24	1	25
25 : Ex 25	2	10
26 : Ex 26	0	3

4.4 Expert Interview Results

Before discussion of the results from the expert interviews it is important to highlight that their experience of responding to the examples was different from that of the survey respondents. The experts responded to the examples in one-on-one interviews with the researcher and the questions and responses were verbal as opposed to written. As explained in Chapter 3 (Methodology), the expert interviews allowed for a more open and interactive discussion. As the experts were approached for their participation due to their specific music industry experience and consented to their responses being recorded, their responses were not anonymous. However, their identities are not reflected in the presentation of their responses. The interview process led to longer responses in which the experts were able to elaborate on their initial feelings of a given example. These interviews supplied the researcher with additional insight that was not logistically available in the participant surveys.

One key aspect of the expert responses that differs from that of the participant surveys is the gender of the expert sample. Table 4.6 shows that, of the eight experts interviewed, only one is female. This was not the intention of the researcher as a more equal distribution between

male and female participants would undoubtably result in clearer, unbiased data. The expert sample size was intended to be larger (12-15). Other female experts were contacted regarding their participation, but due to unavailability or reluctance they did not participate in the interview process.

Table 4. 6: Expert Respondents' Gender

Gender	**Number of experts per gender**
Male	7
Female	1

Once the expert interview transcription process was completed the coding revealed a similarity with that of the surveys. The themes that were derived using bottom-up coding resulted in the same coding groups found in the participant surveys (see table 4.4). The similarity of the codes used to analyse both the participant surveys and the expert interviews is likely due to the researcher's own subjective interpretation of the responses. Considering the experts' responses, the researcher could apply the already conceived codes to the new dataset. This illustrates that even though the experts were presented with the examples in a different context, were subjected to a different mode of questioning, and were able to discuss and elaborate their responses, they still responded in a similar way to the survey participants.

Table 4. 7: Expert Respondents and All Codes

	A : Anthropological Description	B : Emotion	C : Imagery	D : Music	E : Other	F : Situations	G : Sonic Description	H : Speech	I : Technology
1 : Ex 1	6	7	2	9	1	0	0	5	1
2 : Ex 2	1	5	1	6	0	0	7	7	1
3 : Ex 3	3	13	0	3	0	1	11	4	2
4 : Ex 4	4	2	1	8	0	0	9	5	0
5 : Ex 5	3	16	3	12	0	3	8	2	0
6 : Ex 6	3	11	3	4	2	1	2	6	1
7 : Ex 7	6	12	3	4	1	1	6	3	1
8 : Ex 8	3	6	5	16	0	2	7	2	1
9 : Ex 9	12	0	1	7	0	4	5	11	0
10 : Ex 10	14	5	4	5	0	0	6	9	1
11 : Ex 11	4	13	3	25	0	2	6	2	1
12 : Ex 12	2	6	3	14	0	0	3	12	0
13 : Ex 13	9	5	1	8	0	2	2	10	1
14 : Ex 14	9	3	2	9	0	0	6	4	0
15 : Ex 15	4	17	2	2	2	3	6	6	0
16 : Ex 16	5	7	1	15	0	3	4	4	0
17 : Ex 17	3	11	2	8	0	1	6	4	3
18 : Ex 18	7	7	3	0	0	2	6	7	2
19 : Ex 19	1	14	6	5	0	1	10	2	3
20 : Ex 20	10	21	2	3	2	1	7	3	0
21 : Ex 21	2	2	4	15	0	2	4	7	3
22 : Ex 22	6	7	0	12	0	2	3	8	3
23 : Ex 23	1	9	11	11	0	1	15	3	1
24 : Ex 24	5	6	1	3	1	1	4	11	6
25 : Ex 25	5	4	1	16	0	4	2	6	2
26 : Ex 26	1	7	8	13	1	1	17	2	0

The experts' reference to music represents a point of discussion. It stands to reason that music industry professionals would be presented with composed audio, whether speech or otherwise, and be compelled to respond using musical terminology. The same analysis could be interpreted from the participant surveys, but because participants' names were not recorded with the responses, it is not empirically evident in the data. Although it is known that of the survey respondents, 53% were university music students, it is not clear that every reference to music was supplied by a music student. However, in the case of the expert responses it is known that each respondent is a music industry professional.

4.5 Likert Scale Questions

This section will discuss the responses to the Likert scale questions, discussing the participant and expert responses briefly before comparing how the two groups responded.

Table 4.8 displays how the participant survey respondents responded to each example when prompted "The example is happy". This table reveals that for most of the examples, participants felt neutral or disagreed with the statement. If we compare the examples in which the emotion

of the initial delivery was happy (7, 9, 10, 11, 16, 17, 20, 22, and 26) with how participants responded, we can see that Examples 7, 16, 17, and 20 “Agree” or “Strongly Agree” received the highest reference count. Examples 9 and 10 received the highest reference count for “Neutral”, and Examples 11 and 26 received the highest reference count for “Disagree” (Example 22 had “Neutral” and “Disagree” both with 19 references). This comparison shows that even though these nine examples were initially delivered in a happy expression, the cut-up process and additional processing that was applied to the voice resulted in only four of the examples retaining any identifiable artefacts of the initial delivery.

Table 4. 8: Participant Survey Respondents' Responses to Happy Likert Question

	A : Strongly Agree (H)	B : Agree (H)	C : Neutral (H)	D : Disagree (H)	E : Strongly Disagree (H)	F : Did not answer (H)
1 : Ex 1	1	6	22	11	5	0
2 : Ex 2	1	10	13	12	9	0
3 : Ex 3	0	3	20	18	2	2
4 : Ex 4	0	2	16	18	6	3
5 : Ex 5	1	14	15	8	7	0
6 : Ex 6	0	2	7	16	20	0
7 : Ex 7	2	17	12	11	2	1
8 : Ex 8	3	11	6	18	7	0
9 : Ex 9	1	4	23	14	2	0
10 : Ex 10	1	10	23	9	2	0
11 : Ex 11	4	7	7	18	8	1
12 : Ex 12	6	21	8	8	2	0
13 : Ex 13	0	4	21	18	2	0
14 : Ex 14	1	6	23	6	7	2
15 : Ex 15	1	3	6	19	16	0
16 : Ex 16	3	21	15	5	1	0
17 : Ex 17	2	17	10	15	1	0
18 : Ex 18	0	3	21	17	4	0
19 : Ex 19	1	2	3	24	14	1
20 : Ex 20	24	11	6	2	2	0
21 : Ex 21	3	10	18	9	5	1
22 : Ex 22	0	3	19	19	2	2
23 : Ex 23	0	2	12	21	10	1
24 : Ex 24	0	6	28	8	3	0
25 : Ex 25	6	20	9	6	2	2
26 : Ex 26	0	7	15	17	5	1

Table 4.9 shows the expert responses to the Happy Likert scale question for each example. Unlike the participant responses, the expert responses show a more even dispersion between “Agree”, “Neutral”, and “Disagree”. Examples 7, 11, 16, 17, 20, and 22 all have “Strongly Agree” or “Agree” as the most common reference and Examples 9 and 10 have equal highest between “Agree” and “Neutral”. Of the nine examples with a happy initial delivery, the experts were able to hear this quality in seven.

Table 4. 9: Expert Respondents' Responses to Happy Likert Question

	A : Strongly Agree	B : Agree	C : Neutral	D : Disagree	E : Strongly Disagree
1 : Ex 1	0	1	5	1	0
2 : Ex 2	0	2	2	3	0
3 : Ex 3	0	2	0	3	2
4 : Ex 4	0	2	3	2	0
5 : Ex 5	2	1	1	3	2
6 : Ex 6	0	1	1	3	3
7 : Ex 7	1	4	1	2	0
8 : Ex 8	1	1	2	2	2
9 : Ex 9	0	4	4	0	0
10 : Ex 10	0	3	3	2	0
11 : Ex 11	0	4	1	3	0
12 : Ex 12	0	5	2	0	1
13 : Ex 13	0	1	5	2	0
14 : Ex 14	1	4	2	1	0
15 : Ex 15	0	0	0	4	4
16 : Ex 16	1	5	1	1	0
17 : Ex 17	0	4	2	1	1
18 : Ex 18	1	1	1	3	2
19 : Ex 19	0	0	2	3	3
20 : Ex 20	7	0	1	0	0
21 : Ex 21	0	1	3	4	0
22 : Ex 22	1	4	2	1	0
23 : Ex 23	0	0	5	3	0
24 : Ex 24	0	2	4	1	1
25 : Ex 25	0	5	2	1	0
26 : Ex 26	0	1	2	3	2

Considering the second research question, when asked if the example was happy, the participant survey data suggests that the cut-up and composition process that was used in the creation of the examples changed the emotive expression of the cut-up speech sounds; while the expert responses suggest a contrasting interpretation as they were able to identify the initial emotive expression for seven out of nine happy examples.

Table 4.10 represents how the survey participants responded when prompted “The example is fast”. The most common responses were “Agree” or “Neutral” with the table illustrating a reluctance to “Disagree”, with the exception of Examples 6, 15, 16, and 23. The researcher’s hypothesis for this tendency relates to the fact that the experience of hearing cut-up sounds (speech or otherwise) may make a receiver feel rushed or interrupted, therefore giving the impression that the sound is fast.

Table 4. 10: Participant Survey Respondents' Responses to Fast Likert Question

	A : Strongly Agree (F)	B : Agree (F)	C : Neutral (F)	D : Disagree (F)	E : Strongly Disagree (F)	F : Did not answer (F)
1 : Ex 1	3	10	15	15	1	1
2 : Ex 2	2	26	8	7	2	0
3 : Ex 3	15	21	6	2	1	0
4 : Ex 4	0	19	17	8	1	0
5 : Ex 5	19	18	4	1	3	0
6 : Ex 6	0	7	15	19	4	0
7 : Ex 7	1	27	12	2	3	0
8 : Ex 8	0	7	22	10	5	1
9 : Ex 9	0	21	13	6	5	0
10 : Ex 10	1	8	22	12	2	0
11 : Ex 11	0	11	24	7	3	0
12 : Ex 12	6	21	15	1	2	0
13 : Ex 13	6	14	18	6	1	0
14 : Ex 14	6	25	10	2	1	1
15 : Ex 15	0	3	18	19	5	0
16 : Ex 16	0	4	18	20	2	1
17 : Ex 17	13	28	2	2	0	0
18 : Ex 18	5	16	20	3	1	0
19 : Ex 19	1	14	15	11	4	0
20 : Ex 20	1	32	8	2	0	2
21 : Ex 21	3	26	10	2	1	1
22 : Ex 22	8	24	11	1	0	1
23 : Ex 23	2	11	11	13	6	1
24 : Ex 24	0	13	21	8	2	1
25 : Ex 25	1	16	19	7	1	1
26 : Ex 26	1	3	18	15	8	0

Table 4.11 represents how the experts responded when prompted “The example is fast”. When compared with Table 4.10 a more even dispersion can be seen in the expert responses across the range of examples. Examples 6, 15, and 16 received the highest reference count for “Disagree” in both groups, but the experts appear more varied in their responses.

Table 4. 11: Expert Respondents' Responses to Fast Likert Question

	A : Strongly Agree	B : Agree	C : Neutral	D : Disagree	E : Strongly Disagree
1 : Ex 1	0	2	1	4	0
2 : Ex 2	0	3	0	4	0
3 : Ex 3	2	4	1	0	0
4 : Ex 4	0	0	1	5	1
5 : Ex 5	3	4	0	0	0
6 : Ex 6	0	0	3	4	1
7 : Ex 7	3	4	1	0	0
8 : Ex 8	0	1	2	4	1
9 : Ex 9	1	2	4	1	0
10 : Ex 10	0	2	2	4	0
11 : Ex 11	0	2	1	5	0
12 : Ex 12	1	3	2	2	0
13 : Ex 13	0	2	4	2	0
14 : Ex 14	1	3	2	2	0
15 : Ex 15	0	1	1	3	3
16 : Ex 16	0	1	2	4	1
17 : Ex 17	2	5	0	1	0
18 : Ex 18	1	5	1	1	0
19 : Ex 19	2	1	4	1	0
20 : Ex 20	1	5	1	1	0
21 : Ex 21	0	3	2	3	0
22 : Ex 22	1	7	0	0	0
23 : Ex 23	0	2	4	1	1
24 : Ex 24	0	1	5	1	1
25 : Ex 25	0	2	3	3	0
26 : Ex 26	1	0	2	3	2

4.6 Phrases

Tables 4.12 and 4.13 show the responses of the participant survey respondents and expert respondents (respectively) when asked whether they could hear a word or phrase in the examples. In the participant survey responses, of the 26 examples, only five examples (3, 14, 18, 20, and 24) have the highest reference count for "Yes". Interestingly, only eight examples show reference counts of ten or lower, which means that for the other eighteen examples, at least 22% of the participants were able to hear words or phrases. The researcher had hypothesised that the looped examples (2, 8, 12, 17, 21, and 25) would result in high reference counts for "Yes" because the repetitive nature of the cut-up speech sounds may allow time for participants to make linguistic sense of the sounds. Therefore, if a cut-up sequence of cut-up speech sounds is repeated or looped within a piece of music, a listener may be able interpret a degree of linguistic information from the fragments. This notion suggests that it may be extremely difficult to completely divorce the emotive expression of speech from its linguistic signification. Examples 2, 8, 17, 21, and 25 received relatively high reference counts for "Yes", but "No" was the common response across the five examples (except for Example 8 among the expert respondents). In fact, the five examples in which the most common reference was "Yes" (3, 14, 18, 20, and 24) are all examples that were not edited using one of the following

techniques: time-stretch, reversing, or looping. These three editing techniques carry the most potential to obscure the natural qualities of the voice. One could conclude that the more the original tone of a voice is altered, the harder it is to perceive linguistic meaning from fragments of speech.

Table 4. 12: Participant Survey Respondents' Responses to Phrase

	A : Yes	B : No	C : No answer
1 : Ex 1	5	39	1
2 : Ex 2	15	30	0
3 : Ex 3	22	22	0
4 : Ex 4	18	27	0
5 : Ex 5	3	41	0
6 : Ex 6	19	25	0
7 : Ex 7	8	37	0
8 : Ex 8	7	37	0
9 : Ex 9	15	30	0
10 : Ex 10	7	38	0
11 : Ex 11	18	27	0
12 : Ex 12	18	27	0
13 : Ex 13	14	31	0
14 : Ex 14	33	12	0
15 : Ex 15	15	30	0
16 : Ex 16	18	27	0
17 : Ex 17	21	24	0
18 : Ex 18	28	16	0
19 : Ex 19	10	35	0
20 : Ex 20	23	22	0
21 : Ex 21	22	23	0
22 : Ex 22	3	42	0
23 : Ex 23	11	34	0
24 : Ex 24	32	13	0
25 : Ex 25	19	26	0
26 : Ex 26	4	41	0

Table 4. 13: Expert Respondents' Responses to Phrase

	A : No	B : Yes
1 : Ex 1	7	0
2 : Ex 2	2	5
3 : Ex 3	3	5
4 : Ex 4	2	5
5 : Ex 5	8	0
6 : Ex 6	2	6
7 : Ex 7	8	0
8 : Ex 8	7	1
9 : Ex 9	4	4
10 : Ex 10	5	3
11 : Ex 11	7	1
12 : Ex 12	2	5
13 : Ex 13	5	3
14 : Ex 14	1	7
15 : Ex 15	3	5
16 : Ex 16	2	6
17 : Ex 17	4	4
18 : Ex 18	1	7
19 : Ex 19	8	0
20 : Ex 20	2	6
21 : Ex 21	1	7
22 : Ex 22	8	0
23 : Ex 23	7	1
24 : Ex 24	1	7
25 : Ex 25	2	6
26 : Ex 26	7	1

However, the expert respondents' results show a complete opposing response. Examples 2, 3, 4, 9*, 12, 14, 15, 16, 17*, 18, 20, 21, 24, and 25 ((*) implies the counts were evenly split) all display the highest reference count for "Yes". The experts were able to hear words or phrases in 14 of the 26 examples. This could potentially be a result of the response method; the experts were able to verbally respond to the questions rather than write out the answers. They may have felt more pressure in a one-on-one interview situation to search for words or phrases.

4.7 Correlation charts

Table 4.14 shows the correlation coefficient between the most common themes referenced by participants and the experts ("Anthropological Descriptions", "Emotion", "Music", "Sonic Descriptions", and "Speech"). The coefficient reads 0.631**. This means that there is a

significant relationship between the participant and expert reference counts for each example. In other words, when the participants responded with high references for theme, the experts responded in a very similar manner.

Table 4. 14: SPSS Spearman's Rank Correlation Coefficient Table for the Most Common Participant Survey and Expert Respondent Thematic Codes

Correlations				
			Most Common Participant Theme	Most Common Expert Theme
Spearman's rho	Most Common Participant Theme	Correlation Coefficient	1,000	.631**
		Sig. (2-tailed)		0,002
		N	26	22
	Most Common Expert Theme	Correlation Coefficient	.631**	1,000
		Sig. (2-tailed)	0,002	
		N	22	22
**. Correlation is significant at the 0.01 level (2-tailed).				

The significant correlation between how the participants and experts responded shows that regardless of industry expertise the experience of the listeners was similar. Part of working with stimulus (examples) that contain so many variables (source; gender; emotional delivery; articulation; processing, etc.) is that when a strong positive or negative correlation is shown between two examples, it can be challenging to accurately conclude the reason for the correlation. To ascertain why two examples displayed significant correlation, each example was given an intelligent label that would represent the details of the sound. [6] Subsequently, correlation analysis was explored for all of the parameters in the intelligent labels to establish the reasons that participants and experts responded as they did. The results of these charts are discussed below.

[6] Appendix: Table A2.9

4.8 Intelligent label charts

Examples with a female source were analysed for correlation to determine if there was any link between how participants and experts responded to examples with a female source. Only four instances of significant correlation occurred out of a possible 45 relationships (8.89%). Examples with a male source show a similar relationship. 14 instances of significant correlation are present (12.5%). This means that there is very little relationship between the responses when analysed only in terms of the gender of the source voice.

Many of the intelligent labels' correlation analysis returned similarly insignificant percentages when compared based only on one parameter of the example. The most noteworthy of the single parameter comparison results were for the examples with tape delay (18.18%), neutral emotion (20%), reverse (20%), tape delay with modulation (30%), time-stretch (30%), Sesotho phrases (33.33%), but the largest was looped examples (53.33%). As previously mentioned, the reasons for these correlations are challenging to establish because the examples contain many variables that could be the reason for the similar trend. However, to establish if any combination of variables influenced the participants and experts to respond in a particular manner, examples were compared based on combinations of parameters and analysed for correlation. The most noteworthy results of these comparisons are outlined in the table below:

Table 4. 15: Percentage of Significant Correlations Between Examples based on Shared Characteristics [7]

Variables	Number of relationships	Number of significant correlations	Percentage of significant correlations
TS + R + L	78	21	26.92%
L (no R or TS)	6	3	50%
R + L (no TS)	45	11	24,44%
GM + L	6	2	30%
GM + R	3	1	30%
GM + TS	6	1	16.67%
All TD	136	31	22.79%

[7] See appendix: Table A2.9

GF + EMH	10	2	20%
GM + EMS	3	1	33.33%
L + TD	6	3	50%
GM + EMN	10	2	20%
All L + TS	28	10	35.71%

Still, the highest percentage of significant correlations was between all looped examples (L). The following table breaks down the correlations of all the looped examples and their similarities.

Table 4. 16: Breakdown of SPSS Spearman's Rank Correlation Coefficient for all Looped Examples

Correlation	Similarities				
0.010	GM	L	EHPNN	TD	
0.194	L	TD/TDM			
0.230	L	TD			
0.345	GM	ARCON	L	TD	
0.385	L	TD/TDM			
0.458	GM	TS	L	TD/TDM	
0.487	GM	ARCON	L	TD/TDM	
0.579*	L	TD			
0.581*	ARSH	EMA	L	TD	
0.613*	L	TDM			
0.620*	L	TD/TDM			
0.680**	GF	L	TD		
0.694**	GM	L	TD		
0.699**	L	TD/TDM			
0.715**	GM	ARCON	EMS	L	TD/TMD

Table 4.16 gives little insight into the potential reasons for correlation between looped examples other than that they are looped. Looping is a common technique in many styles of music involving multiple repetitions of an idea. This can take the form of individual elements of a musical arrangement, called ostinatos (drum patterns; chord progressions; melodies; lyrics;

etc.) or entire sections of a musical work. Looping is a form of repetition and repetition is strongly related to rhythm (when something happens in time). Rhythm and repetition are common phenomena in all of life; so prevalent that our attention is easily drawn to them. The compositional analysis of figures 6.2 and 6.3 reveals that rhythm (specifically repetition) is a key element in perception of sound. The use of rhythm in music can alienate and disorientate listeners, but just as easily facilitate shared experience.

5. Exploring the Data in Compositional Practice

The initial intention of the researcher to create a musical taxonomy of cut-up speech sounds that could be used to communicate emotional expression in their music and aid in musical accessibility is considered here. This chapter examines the challenges and benefits of using qualitative data to inform musical composition from the researcher's own experience of creative practice experience, and describes how the outcomes of the research led to a change in compositional approach. The chapter concludes with the researcher's experience of utilising data to inform creative practice and discusses suggestions for further explorations of this topic.

The rationale for creating the taxonomy was to signify meaning through the researcher's musical practice, with an understanding of the associative relationships that an audience would have with a particular method of delivering cut-up speech sounds. Upon reviewing the results of the participant surveys and expert interviews, the researcher was able to see one characteristic in the data with relative ease: across the responses, there was a wide range of variation. Although the responses to certain examples revealed that there was some shared interpretation of the sounds, a large majority of the references were diverse. For example, 50 references to "Music" were recorded for Example 16, the highest single reference count of any code for the participant survey. However, for the same example, 47 references were split between "Anthropological Descriptions"; "Emotions"; "Imagery"; "Sonic Descriptions"; "Speech"; or "Technology". These kinds of responses illustrate that the experience of each participant varied considerably and that systematically classifying the sounds used in the creation of the examples would be based on weak correlations in the data, rather than unanimous agreement.

Following this realisation, the focus shifted away from creating a taxonomy of each cut-up speech sound, to a more interpretative reading of the results that could aid the timbral arrangement of the cut-up speech sounds in an equally emotive way. The data certainly derived rich insight regarding the participants' responses and was used to understand how a listener might respond to sounds, production techniques, phrase arrangement, etc. in the musical works. When compared with the characteristics of the examples' initial delivery, the way it was edited, and how it was presented, the results function as another chapter in the narrative of the composition process. As the arbiter of each step of the data gathering process: conception of

the idea; recording engineer for the source material; composer of the examples; designer of the questionnaire; interviewer; the researcher could reflect on each step of the process and use their insight to try and make sense of the reasons behind the participant results. This insight is subjective in nature and clouded by time and memory, however, it proved to be essential knowledge when engaged in the creative practice of exploring the musical applications and potential significance of the cut-up speech sounds. For example, Source A: phrase 1 was delivered with a sad emotive expression. When presented to the participant survey respondents, the most common emotive description of the example was "Fear". [8] This change demonstrates that the emotive expression of the cut-up speech sounds was altered through the cut-up and processing of the voice and expands the possibilities for musical application of that cut-up speech sound. The thematic coding and correlation analysis were conducted to try and determine the reasons for this kind of change in the experience of the voice, but the reasons for this change are vague. Had the analysis illustrated a clear reason for the responses, then a taxonomical approach could have been employed to organise the cut-up speech sounds in a musical context.

Compositionally, the fact that the perception of emotive expression seems to change based on the contextual arrangement, gave the researcher liberty to consider the timbral and "physical" qualities of the cut-up speech sounds as potential creative avenues. The knowledge that one cannot accurately predict a listener's emotive response to music based on the cut-up speech sounds presented an opportunity for the researcher to create a rich musical context that could guide a listener through their own subjective experience of the music.

The cut-up speech sounds of Source A: phrase 1 were commonly utilised by the composer through the album. Tracks such as "Midnight's Children", "The Weary Traveller", "Undesirable", "Day Drifter", "Melancholy Folly", and "Do You Understand?" each make relatively overt use of Source A: phrase 1 but the application of the cut-up speech sounds and the emotive expression of the sounds is different depending on the musical context of each track. The varied emotive expression that the cut-up speech sounds from Source A were able to convey in each track would not have been possible if approached using a taxonomical methodology in which each cut-up speech sound was attributed an emotional signification.

[8] See Appendix: Table A2.2

As discussed in the previous chapter, the participant and expert responses to the emotive expression of the examples spoke to the second research question. In relation to the third research question, the researcher's experience of the creative practice suggests that it would be challenging for a composer to extract specific cut-up speech sounds from a spoken phrase, utilise the sounds in a musical work, and have the emotive expression remain unchanged without creating a subjective conceptual space that could aid or suggest to a listener the same emotive experience. Following this, with the knowledge of musical accessibility being highly dependent on the cultural context and subjective experience of a listener, the emotive expression of the music would be challenging to predict.

The disadvantages or challenges of using data to inform the musical materials were such that the data may confine or limit the way one perceives the cut-up speech sounds. 39 references were made to "Joy" for Example 20 and 33 references were made to "Fear" for Example 6. This implies that the sounds invoked certain responses from the listeners and that those same or similar sounds should be used in a musical context to invoke the same emotion.[9] The reason for this response may be linked to a number of variables in the examples, ranging from the original emotive content of the phrase, the manner in which the sound was cut-up, the effects placed on the audio, or any combination of these variables. Furthermore, the association that the data reflects for a specific example may be at odds with how the researcher interprets the sound. The results of the happy Likert scale questions show that the context of the cut-up speech sounds change the emotional expression of the original message. For example, the original emotional delivery of Examples 7, 9, 10, 11, 16, 17, 20, 22, and 26 was happy. Only Examples 7, 16, 17, and 20 received the most references to "Agree" or "Strongly Agree", while Examples 9, 10, and 22 received high "Neutral" responses and Examples 11, 22, and 17 received high "Disagree" responses. Whether obscured by the cut-up process or the effects placed on the sound, the process of creating the examples changed the way the speech expressed emotion, speaking to the second research question.

However, there were advantages to using the data to inform the musical works. Any insight into the nature of the raw materials of a creative process can help a practitioner construct meaningful art. The knowledge that consonant and dissonant intervallic relationships between notes derived from the harmonic spectrum are used to construct various harmonic tonalities

[9] See Appendix: Table A2.2

and that one's association with these tonalities informs how they are used to convey musical meaning, is a creative tool for musicians. Similarly, the researcher's insight into the participants' experiences, and his own experience, has been a process in understanding the broad way in which the music will be received. The analogy of a recipe seems apt here. A dish may contain any number of ingredients, flavours, and spices, which on their own are experienced as sweet, savoury, spicy, bitter, etc., but depending on how the ingredients are prepared, combined, cooked, and presented, the eating experience can vary drastically. A chef uses their knowledge of the ingredients and cooking methods to create varied flavour combinations. In much the same way the results of this research supplied more knowledge about the materials used to create the music (ingredients). The way these materials have been treated was influenced by personal taste, intuition, and experimentation to create interesting combinations of cut-up speech sounds. The more materials added and treated in the process obscure the clarity of the original ingredients but result in depth of flavour and complexity that are exciting as an artist.

When engaging with examples that sparked inspiration, the insight gleaned from the responses was used as a comparison for the researcher's own feelings and association with the sound. If while "mining" the various examples for viable material, the researcher was left uninspired or without a clear method to treat the sounds, the responses of the participants were used to guide the creative process from the outset. "The next piece should feel…"; this question was considered, and the data was examined to inform the use of cut-up speech sounds for the researcher to use in construction of the piece. For example, the researcher would review the responses and see that in response to Example 24, 32 participants were able to hear a word or phrase. Out of this knowledge the concept of "the next piece will play with the idea of a phrase, obscured, and scattered at first, revealing itself over time" was developed. Often the initial idea for a piece was the hardest to come by, and in these cases the results supplied the researcher with themes and material with which to begin a new piece, create an Ableton Live software instrument, edit sounds, or clarify the focus of existing thematic material.

Finally, for the researcher to have been able to create a taxonomy of the cut-up speech sounds based on participant and expert responses, more detailed information would be necessary. The level of insight required to understand the reasons for a participant's response would require a far more in-depth questioning process, in which a participant would have more time to elaborate on their associations and compare their responses with unedited control versions of

the same speech. This assumes that a participant would be able to conceptualise why they responded in a certain way. The reasons for their associations may not appear clear to them even with the luxury of more time and the ability to compare. A logical subsequent question also needs to be considered in relation to a musical taxonomy. If a taxonomy were to be created, based on more detailed data, would music using such a taxonomy express meaning in a predictable manner? For such a taxonomy of cut-up speech sounds to produce an analogous experience in a listener, the music created may have to be played to the same participants that supplied the responses used to create taxonomy. Outside of this new hypothetical sample, the experience of a listener is too complex and subjective for a predictable and wider shared experience to be certain.

One of the primary research questions asked if a composer can extract and harness the speech sounds which allow one to signify emotive meaning through spoken language. The researcher believes the answer is both yes and no. Yes, if it is possible to edit the audio in such a way that only the emotive meaning of a phrase remains, a composer may be able to utilise the cut-up speech sounds in a musical taxonomy to convey the original emotional content. However, outside of their own experience of the cut-up speech sounds, it would be challenging to predict a listener's experience.

6. Presentation, Analysis, and Commentary of Music

The styles and discourse of popular music were used as the aesthetic influence for this researcher's music as it relies on established forms and production techniques recognisable to a general listener. As a result, the researcher could utilise and manipulate these forms and techniques to experiment with the emotive expression of cut-up speech sounds in a musical context. Thus, presenting novel conceptual ideas that are grounded in accessible musical forms. The artistic concept that governs the album is that each track is a metaphysical exploration of the researcher's experience of conversing with various fictional characters. Each track presents a musical representation of the researcher's sentient experience of a character's intentions, mannerisms, emotions, and expression. The concept was informed by: Susanne Langer's description of music's ability to express meaning as a reflection of sentient experience (Langer 1942: 222, 223, 228) and the researcher's own sense from early in the research process that the cut-up speech sounds felt analogous to voices in one's head. This idea was influenced by the book *Midnight's Children* (1981) by Salman Rushdie in which the protagonist Saleem is born with the gift of hearing the thoughts of others as voices in his head. These voices, fragmented and confusing, felt as though they had information to communicate, but the exact nature of the information was not clear, resulting in a vague sense of meaning.

The tracks largely conform to popular song formal structure: verse (A) and chorus (B) material that utilises common time signatures. As stated in Chapter 1 the researcher's artistic intention was to present novel, and at times challenging, musical material in an accessible manner. One way in which the music embodies this idea is through its adoption of an unambiguous, popular music formal structure, with an emphasis on the timbral quality of the cut-up speech sounds and the manipulation thereof. When presented in this manner, the music employs an accessible contextual base from which a listener can explore the material.

6.1 Midnight's Children

Directly referencing the eponymous novel and its influence on the conceptual framework of the album, "Midnight's Children" acts an aesthetic introduction of the album. The aim is for the listener to imagine themselves as the protagonist experiencing a newfound ability to perceive the inner voices of others.

Table 6. 1: "Midnight's Children" Formal Structure

Form	Intro	A	B	A	C	A	C	B	Outro
Length (Bars)	8	16	20	16	8	16	8	16	10
Tempo	143 bpm								
Time Sig.	12/8 & 4/4								
Duration	3' 21''								
Instrumentation	Drums & Source A (Phrases 1, 2, 4, and 5)								

The researcher created an instrument rack in Ableton Live that would interface with a sample pad controller. Each pad controlled a range of cut-up speech sounds with similar timbral characteristics from a particular source. The pads sequenced through the available sounds each time they were struck, meaning that a player could affect the rhythm and velocity with which the sounds were played while limited to the pre-set order of the sounds. Initially, the instrument rack would select a random sound from those programmed to each pad, but the result was far too varied for the purposes of this track, the intention of which was to create a groove-based environment in which the interplay between the cut-up speech sounds would act as the focal point of the music, while supporting the groove and rhythmic accompaniment.

The initial iteration of this track was one of nine improvisations created using the above-mentioned triggering method. All nine improvisations were recorded on the 27th of March 2023. The source material for this track was percussive sounding cut-up speech sounds taken from Source A and Source C (see Table 3.3). The tones of the two voices contrasted in a way that made them audible, while exhibiting the percussive quality needed to work with the drums. The drums and the MIDI input were recorded simultaneously.

The result of this initial improvisation was a piece lasting 2m 17s. Structurally the piece was lacking in large-scale form that would present material, rework, and contrast the thematic content, but the researcher felt that it succeeded in establishing a strong, engaging, and accessible groove environment that could be further developed. The thematic content that resulted was influenced by shuffle drum patterns in which the drummer plays an ostinato of the first and third triplet partials of each beat, which has the effect of implying forward motion and energy. This was played by the right hand moving between the snare drum (muted by the palm of the left hand with the snare wires turned off) and the two sample pads. The left hand played a supporting role to the shuffle pattern in the right hand by filling in the second triplet

partials of some of the beats. When the hand pattern moved away from the shuffle rhythm, a four note paradiddle phrasing was played to give a four-against-three polyrhythmic feeling to the groove. A regular quarter note pulse was played on the bass drum with the high-hat playing a back-beat (beats 2 and 4). This ostinato is commonly found in swing drumming. In a swing context the technique of lightly playing the bass drum on all four quarter notes is called feathering. In this piece, however, the bass drum was played with greater emphasis to create a strong foundational pulse like that found in EDM and disco.

The corrections needed for this piece applied to the formal organisation: sections were needed so that the thematic content could be presented, developed, and then contrasted to avoid redundant repetitions.

The material was reviewed for natural breaks in phrasing that resulted from the improvisation. From these breaks, sections were outlined and organised into a basic song form (verse – chorus – verse – chorus – bridge – chorus). Much of the thematic content was kept from this stage of the process. The resulting piece followed a basic large-scale song structure but lacked elements expected from a popular song. These were lead (focal element or idea), and pad (background/supporting harmonic content). The lead was created using more melodic cut-up speech sounds from Source A. To contrast the busy foundational rhythmic content, the lead line in the verses consisted of short triplet-based ascending melodic phrases with rests to emphasise the beginning and end of the phrases. The choruses employ a call and response style melodic structure between the lead voice and supporting rhythmic phrases (also Source A). The bridge section breaks away from the triplet feel of the verse and chorus. A straight sixteenth note feel was used to give the feeling of speeding up. Two-bar phrases were repeated four times to increase tension and allow the piece to feel as though it opens up when returning to the verse and chorus material. A pitched-down and time-stretched speech sound from the lead voice was used to act as a pad for each section. The pad is further pitch-shifted to outline a sense of harmony. Due to its low-pitched nature, the pad also functions as a bass line throughout the song. With all the changes to the arrangement and thematic content from the initial improvisation, the drum accompaniment no longer worked to support the rest of the elements. The drums were re-recorded to better support the lead and supporting rhythmic elements in each section. The shuffle pattern remains but functions more as an embellishment of the lead line. The way that time is felt is in half-time which, in conjunction with the shuffle, gives the music a sense of rising and falling that emphasises the call and response in the lead melody.

6.2 The Weary Traveller

"The Weary Traveller" presents the listener with a tired and slightly disheartened character that has great depth and understanding (is organised, structured, and confident in their thoughts) but who also displays longing and sadness, stemming from the knowledge that the information gained from years of searching for meaning has not given them purpose.

Table 6. 2: "The Weary Traveller" Formal Structure

Form	A	B	A	B	C	A	B	A	B	A
Length (Bars)	26	16	18	8	19	2	2	8	10	16
Tempo	90 bpm									
Time Sig.	4/4									
Duration	5' 36''									
Instrumentation	Drums, Source A, B, C, D, E									

This thematic content developed out of frustration with the process of improvisation and generating material in a "live" setting. The researcher felt limited by his ability to improvise engaging and aesthetically pleasing material and wanted to explore the process of developing a piece from several brief melodic ideas. A technique used to develop freedom and control when practicing the drums is to play an ostinato with one or more of one's limbs and use reading pages (monophonic rhythmic phrases) as the solo or melody. The researcher was inspired by this process and created several cells or one bar melodic ideas that could be arranged into a longer form piece.

A lead voice was chosen from the five sources and acts as the melodic focus in each section. Section A uses Source A as the lead, section B uses Source E, and section C used Sources B and D. The desire to have a different voice lead each section was to contrast and signal the change from one section to the next. The drums follow the lead voices and change the feel or environment from section to section. In section A the drums are quite syncopated and follow and embellish Source A. The effect results in a sparse groove without losing the perception of pulse. Section B sees the drums breaking away from strictly following the rhythm of the lead voice in favour of a more open and driving sound inspired by drum and bass breaks. Section C moves drastically away from the established feel and thematic content of the previous two sections, opting for a groove environment in which the drums are primarily playing improvised melodic ideas on the tom toms over a regular back beat ostinato on the high hat.

The initial iteration of the piece successfully showed proof of concept. The idea of arranging short melodic cells resulted in interesting and engaging thematic material. Through a combination of short musical ideas, a complex, multi-timbre melodic accompaniment was generated in which an attentive listener could identify repeated patterns, which bore aesthetic appeal. The constant change between the timbres of the different voices in the lead role however, resulted in a sense of disconnectedness.

The first version of this piece did not consider the function of each element beyond lead and accompaniment, so development focussed on balancing each element so that its function was clear. One of the pitfalls of using such short cells to construct melodic information is the repetitive nature and lack of phrasing breaks. The rhythmic and percussive feel of the melodic content meant that although the material was repeated it was possible that a listener many be unable to latch onto the content of the lead voice.

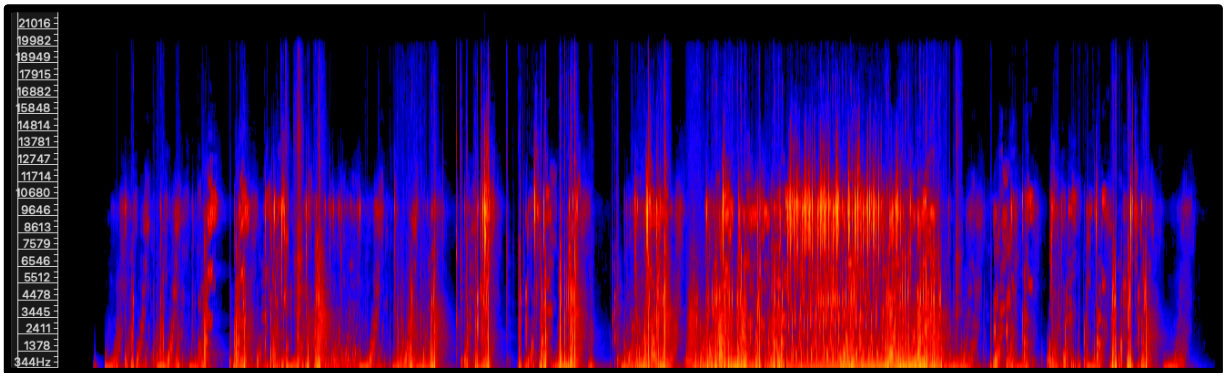
To rectify the perceived shortcomings of the initial piece the researcher composed lead lines that would occupy the forefront in each section. As was initially intended, section A's lead was composed using cut-up speech sounds from Source A and section B's lead was composed from Source E. The speech sounds from both sources were time-stretched and pitched. They utilised longer note values than those of the melodic cells to clearly separate the new lead lines from the melodic cells, which now acted primarily as melodic and rhythmic accompaniment. Section C's focus, in contrast to the focus of the other two sections, places focus on the drums with the voices playing the role of rhythmic accompaniment and pad. The drums remained unchanged.

6.3 Confusion Kills

Table 6. 3: "Confusion Kills" Formal Structure

Form	Through-composed improvisation
Tempo	Free
Time Sig.	Free
Duration	4' 54''
Instrumentation	Drums & Source A, B, C, D, E

Figure 6.3. 1: "Confusion Kills" Spectrograph



Spectrographs have been included for the through-composed pieces ("Confusion Kills" and "Inner Piece" (track 4)) to supply the reader with a visual representation of the development of the music, albeit only representative of three variables (time on the X axis, frequency on the Y axis and the relative energy of frequencies shown as colour shading). The descriptions of the other pieces include colour-coded representations of the formal structure of the music, which are not representative of a through-composed form.

This piece is the result of the same process that gave rise to "Inner Piece" (which follows directly after this track). The two pieces can be thought of as a part-one and part-two. The interaction that takes place in this piece is between three male voices (Sources B, C, and D) and two female voices (Sources A and E). All the source voices were used (Sources A, B, C, D, E).

The mood of this piece differs from "Inner Piece". It feels more ominous and chaotic, with more voices vying for the attention of the listener. One of the male voices (Source B) can be heard repeatedly uttering a pitched down whisper that occupies the low range, along with a time-stretched voice that sounds more like a growl than a spoken utterance. The resulting effect of the voices interacting feels more like an internal struggle than in "Inner Piece" which gives the impression that one is observing a conversation or discussion. This piece feels visceral, as though the listener is directly involved in the process.

Much like "Inner Piece" the only reworkings that took place here were the addition of processing to aid in the creation of a space/context for the voices and the drums. A low frequency delay creates the effect that part of the sound is happening under water, which contributes to the visceral nature of the piece. The result is the feeling that the voices come in

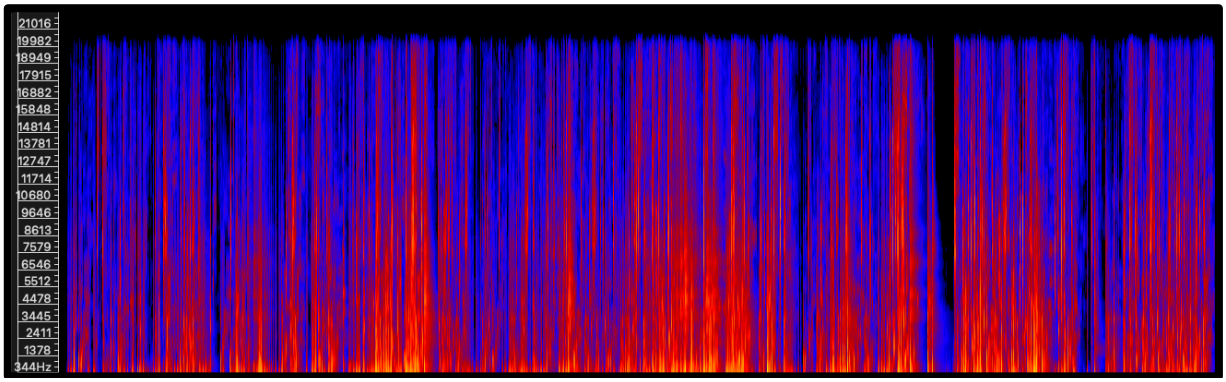
waves, more chaotic and denser with each swell. The climax of the piece comes after roughly 3m 30s and is the result of a prolonged wave of voices that doesn't let-up, supported and propelled by a driving quarter note bass drum ostinato.

6.4 Inner Piece

Table 6. 4: "Inner Piece" Formal Structure

Form	Through-composed improvisation
Tempo	Free
Time Sig.	Free
Duration	3' 52''
Instrumentation	Drums & Source A, B, C, D, E

Figure 6.4. 1: "Inner Piece" Spectrograph



This piece started as an experimental process designed to explore how the researcher could interface with the drums and the Ableton Live session that housed and facilitated playback of the cut-up speech sounds in a real-time feedback loop. The initial conception for the musical material was to take the form of long-form pieces with the drums and cut-up speech sounds working in tandem to create the feeling of conversations, with the drums responding to thematic material in the cut-up speech sounds and vice versa.

Recorded in February 2023 the initial piece makes use of an Ableton Live instrument rack that allows the researcher to randomly trigger speech samples from a preselected set of options. Each pad focuses on a particular source. The interaction that occurs between the drums and the voices is intended to represent the flow of a discussion between three voices with the drums behaving as the narrator or observer witnessing and describing the unfolding events.

The result of the improvisation is an organic, through-composed piece. It begins with one female voice clearly heard with the drums providing context and contrast. As the piece progresses a male voice enters followed by a second, pitched much lower than the other two voices. As the speech phrases get busier, the drums follow the lead of the voices, and the elements work together to outline the natural ebb and flow.

This piece required little in the way of thematic corrections as the intention remained to create an organic and unscripted representation of a moment of interaction, which the researcher felt was captured in the performance. One should consider the interactions we have that are unforeseen and spontaneous where there is no real preparation or fixed outcome required. These conversations are of interest to the composer because of their ubiquity in our lives. Many of our social interactions follow this form and an expression of this phenomena was desired of this piece.

As far as changes to the initial material are concerned, the researcher applied processing techniques to the audio of the voices and the drums. The material lacked a sense of space and context: no interactions are without context, so the voices were treated with delay and reverb to create the illusion of a space in which the interaction took place. The drums were similarly treated to make them feel part of the same environment. The role of a pad was filled by the delayed sound which adds depth to the sonic environment, while being informed by the focal elements of the piece. The thematic content forms the context of the discussion in an iterative way.

6.5 Undesirable

"Undesirable" refences personality traits or habits that one adopts as coping mechanisms. The effect of these habits is clear but reliance on them prevents their being dropped or forgotten. The mood is dark and ominous, but the thematic material is presented in a familiar fashion to represent the way in which we are, consciously or unconsciously, familiar with our harmful or undesirable qualities.

Table 6. 5: "Undesirable" Formal Structure

Form	Intro	A	B	A	B
Length (Bars)	16	13	16	8	15
Tempo	77 bpm				

Time Sig.	4/4
Duration	3’ 32’’
Instrumentation	Drums & Source A, B, C, D, E

This researcher created a piece around a repetitive melodic idea that simultaneously functions as the foundational and lead elements of the piece. A Korg Volca Sample synthesizer was used to pitch down, and time-stretch a vowel speech sound. The result intentionally obscures the characteristics of the voice and mimics the timbre of a synth bass. A two-bar single note bassline was created which repeats for the duration of the piece. To create a sense of development and subvert the expectations of a listener, the bassline was edited so that the phrase is rarely heard in its entirety until half-way through the piece when the drums enter. Here they create a groove-based context for the melody. Additional melodic phrases were added by speeding up and slowing down sequences created on the Korg Volca Sample. Effect sounds were created by time-stretching single cut-up speech sounds and pitching them up by three or four octaves. These sounds were used as fills to separate phrases and as interactive material for the bassline.

The piece was sparse and slow, but timbrally rich. The initial idea to have a repetitive bassline lead the piece meant that the melodic material lacked a sense of harmonic progression, commonly outlined by the bassline.

The bassline needed to be edited to reflect melodic and harmonic development which would propel the music forward. The initial duration was too short and did not take the time needed to develop the thematic material.

A contrasting bassline was created for variation in each section using the original bassline speech sound. The timbres throughout the piece are consistent and the structural elements stay the same, however the thematic content in the bassline and the drums (additional drums were recorded for the new bassline that followed the rhythmic content closely) varies between each section. The form starts with an expanded introduction that introduces the listener to all the sounds that they will encounter as the piece progresses. The introduction does not contain drums so that when they enter, they come as a surprise and contextualise the cut-up speech sounds. The drums are intended to ground the listener as the sparseness and variety of cut-up speech sounds could feel disorientating on their own. The piece moves between two sections in an ABAB form. The A section contrasts the sparseness of the introduction with a melodic

bassline that makes use of a faster harmonic rhythm. The drums and bassline take the lead of the piece supported by pad, rhythm, and fills that enrich the timbral context. The B section sees a variation of the introduction's bassline, which repeats more regularly, take the sole leading role. The drums move into a more supporting role, while still providing a foundation that derives its rhythmic material from the bassline. The musical aesthetic of this track is influenced by trip hop artists such as Massive Attack and Portishead.

6.6 Day Drifter

"Day Drifter" is a representation of the ethereal fantasies one experiences when day dreaming.

Table 6. 6: "Day Drifter" Formal Structure

Form	Intro	A	B	A'	B'	Outro
Length (Bars)	16	16	16	16	16	8
Tempo	140 bpm					
Time Sig.	4/4					
Duration	5' 22''					
Instrumentation	Drums & Sources A, D, E					

A less-explored idea in previous experiments was a chordal homophonic texture in which the chords made use of multiple versions of the same speech sound.

The researcher loaded the cut-up speech sounds from all the sources into a Korg Volca Sample. This instrument allows one to sequence samples in a looped fashion. A sequence was composed using various sounds from the sources (pitch shifted using the controls on the instrument). The sequence was recorded into Ableton Live in a piece-by-piece fashion, meaning that each sound of the sequence was added one at a time until, by the end of the piece, the whole sequence could be heard in its entirety. Once recorded, each sound of the sequence was duplicated so that five layers of each sound would be heard each time. Each of these five layers was treated like voices in a chord structure, outlining separate notes of a chord. Time-stretching was used on layers of the sound to create longer note values and create a richer texture for each utterance of the chords. The result was a through-composed piece that felt as though it unfolded as the piece progressed.

The ideas of harmonic progression and voice leading were not initially considered by the researcher so, although the chords varied and the layers reacted to each other with each new

sound, a sense of development and forward momentum was required to make sense of the chords.

A four-chord progression was composed for the first chord in the sequence based on more standard rules of classical harmony. The material was organised into a theme and variation form beginning with an introduction that contained only the first chord of the sequence and outlined the harmonic movement of the piece. This harmonic progression continues throughout the duration of the piece and each sound in the sequence was edited in pitch so that the harmonic relationship between chords containing the same sounds moved according to the same four-chord harmonic relationship. The version of the sequence in each section is presented four times with the chords of the first repeat all comprising the same intervallic relationships of the first chord in the progression (regardless of source sound), and so on for the second, third, and fourth repeats. The aesthetic influence for this track is Hans Zimmer's film score for *Blade Runner: 2049* (2017). The piece conveys a timbral environment which implies a departure from reality.

6.7 Oblivious

"Oblivious" was so titled because of the way the drums' thematic content was created: oblivious of any other sound or instrumentation. The track's protagonist, voiced by the drums, is not in dialogue with the cut-up speech sounds. Whereas "Stubborn's" (the following track) lead continues despite the surrounding world, "Oblivious" fails to notice its effect on the musical context.

Table 6. 7: "Oblivious" Formal Structure

Form	Intro	A	B	C	D	A	B	C	Outro
Length (Bars)	10	24	4	15	6	18	4	15	8
Tempo	104 bpm								
Time Sig.	Mixed								
Duration	4' 9''								
Instrumentation	Drums & Source A								

The idea for this piece was to have the drums behave as the lead instrument with the cut-up speech sounds acting as the accompaniment throughout, switching the stereotypical roles of these two timbres.

The drums were recorded from an improvisation without cut-up speech sounds. This improvisation outlined sections in an intentional manner, changing the instrumentation used on the drum set as well as the style of playing to create contrast in each section. The researcher then selected various speech utterances from Source A and composed phrases that followed the drums. The sounds used mimic the timbre of the toms on a drumkit because of their sustained pitch and act as harmonic accompaniment for the drums. The cut-up speech sounds largely followed the bass and snare drum patterns of the drums in each section. Sounds that the researcher felt complimented the timbre of the bass and snare drums followed the rhythms directly, creating a kind of harmonic relationship between the two sounds. Time-stretching was used on the cut-up speech sounds of Source A for variety and variation in the pitch of utterances that were repeated multiple times in short succession.

The researcher enjoyed the resulting piece, but it was brief, lasting only 1'50'' or so. The thematic content required more development and reworking to hold the attention of a listener over a longer duration.

The inherent breaks and contrasts in sections from the improvisation allowed the researcher to create a longer form by repeating thematic material and creating reprises of previous sections, something that was not present in the first version of the piece. Cut-up speech sounds from Source D were used as further rhythmic accompaniment and added some contrast to the melodic material of Source A.

6.8 Melancholy Folly

This piece was conceived as the representation of a person caught between childhood and adulthood. The rigidity of the form contrasts the playful and upbeat nature of the thematic content to create the sensation of restriction. The character represented by the piece is having to grapple with two contrasting desires: what is the appropriate dialogue between the two worlds? To grow and develop a blend and acceptance of the two worlds is required. A rejection of either one threatens stagnation.

Table 6. 8: "Melancholy Folly" Formal Structure

Form	Intro	A	A'	B	C	A'	D	C	A'
Length (Bars)	4	16	26	18	16	10	10	16	10
Tempo	149 bpm								

Time Sig.	4/4
Duration	3’ 26’’
Instrumentation	Drums & Sources A, B, C

The initial material for this piece came from a session working on ideas with the sample pad and a single speech sound. The researcher felt that a single speech sound produced less-jarring melodic content.

Using the sample pad to produce melodic content produced a melody that still felt deeply rooted in drumming vocabulary. The result was a “catchy” four bar melody that looped in much the same way as a pop drum groove. Drums were improvised in a drum and bass style with a focus on breaking up the monotony of the melody. Even with the addition of drums, the result was heavily repetitive and did not develop the thematic content.

Although the researcher was not satisfied with the result of the first iteration of this piece, there were elements of the piece that sparked excitement. If treated with care, the idea of catchy melodic material could be an engaging element for the listener to follow as the piece unfolded.

The first development needed was to rework and develop the initial melodic content throughout the piece. This would allow for different sections and create contrast and a sense of climax. The melody was edited to have breaks in-between phrases that would allow a listener to hear each phrase as a single idea or statement, calling or responding in the manner of antecedent and consequent phrases.

The form of the piece borrows from song form and binary form. The first half of the song sets up the main melodic material through an extended introduction. Four bars of solo lead voice are followed by two eight-bar sections that develop the melody and introduce the drums with a repetitive back beat on a cymbal stack to mimic the effect of a hand clap. In the second eight-bar section, an open hi-hat bark is played on the last upbeat of every two bars, giving the rhythmic accompaniment a sense of building, forward momentum. The verse follows with three repeats of an eight-bar loop comprising two four-bar periods. Two bars are placed before the end of the verse to build tension and delay the arriving chorus. The chorus shifts the range of the melody into the higher octaves which, after 1m 14s, gives the piece the feeling of opening up and releasing the tension built during the introduction and verse. The chorus lasts eight bars

and consists of two repeats of a new four-bar period. Subsequently, the second half of the binary form begins with a sixteen-bar thematically contrasting chordal section. The back beat in the drums is replaced with irregular accents giving the impression that the sense of time has slowed down. This section is followed by a reprise of the introduction before moving into a varied, half-time version of the verse that lasts ten bars. This precedes a repeat of the chordal material (post chorus) before ending with a final call back to the introduction.

Although the final version of this track utilises repetitive melodic ideas and song structure, the move away from a single melodic phrase towards a track that focuses on developing and contrasting thematic content over the duration, creates a more engaging and stimulating musical experience.

6.9 Stubborn

"Stubborn" deals with a character that is oblivious to external input and does not grow or develop when opportunities present themselves. The researcher's impression of the lead male voice (Source B) is one of false, or exaggerated, confidence. The lead melodic content is steady and rigid to portray the idea of control. At no point is a risk taken by the character that may result in the music feeling unstable. The various sections present opportunities for the lead to expand or explore new possibilities, but the lead remains closed off and only conforms to the changing form as much as required to keep the track going.

Table 6. 9: "Stubborn" Formal Structure

Form	Intro	A	B	A'	B'	C	B'	C'	D	A'
Length (Bars)	8	8	5	7	8	4	8	16	8	4
Tempo	60 bpm									
Time Sig.	4/4									
Duration	5' 12''									
Instrumentation	Drums & Source B									

This piece was the result of composing material for a pre-determined form. The form (Table 6.9) was conceived to give the track macro-level organisation from the outset and was established intuitively from the researcher's own experience of listening to popular music. It follows a verse-chorus approach to organising the sections. Throughout the revisions of the piece, the initial form persisted.

The form was conceived and marked out in Logic Pro and material was composed to fit each section. Multiple layers were composed with the intention of creating an a cappella-style arrangement in which the voices would fill the roles of a common five-piece pop arrangement. The initial piece did not progress much further than this arrangement.
The first version of the piece was largely uninspiring, but when revisited in search of material for the final album, the researcher began playing thirty-second note ideas using a single speech sound and an arpeggiator. The effect was similar to the kind of low-pitched picking found in metal breakdowns and the basslines of bands like Nerve, "Tetrastigm" (2011), Heernt, "Locked in a Basement" (2006), and the Mark Guilliana Jazz Quartet, "Inter-Are" (2017).

Cut-up speech sounds from Source B were used to create looping syncopated grooves for each of the pre-existing sections. The drums work in conjunction with the loops, occasionally directly following the vocal rhythms and in other cases they function as a supporting accompaniment so that the loops can take a clearer lead role. The original material was completely scrapped in favour of the new groove-based loops.

This piece was inspired by Mark Guilliana's work with Heernt's *Locked in a Basement* and *Beat Music* albums and JoJo Mayer's group Nerve.

6.10 Do You Understand?

The artistic intention was to present two characters that are embodied by a female lead voice and a male lead voice respectively. The experience of the female voice feels to the researcher to be the perspective of the listener. The female voice is acknowledging the confusion and pain of the male voice and trying to empathise. The male voice is defensive and dismissive, often interrupting the female voice's phrasing. The track does not resolve this issue. The two voices are still disconnected by the end of the track, but a conversation has occurred which sheds light on the perspectives of each character (something that is not present from the outset of the track).

Table 6. 10: "Do You Understand?" Formal Structure

Form	A	B	C	A	D	C	D	E	C	B	C	B	C	A	D	C	D	E	C
Length (Bars)	24	18	2	6	8	2	10	14	15	7	1	9	2	18	8	3	8	11	16
Tempo	90 bpm																		
Time Sig.	4/4																		
Duration	8' 17''																		

Instrumentation	Drums, Source A & Source D

Where pieces like “Confusion Kills” (track 3) and “Inner Piece” (track 4) intend to represent the interaction of people in the context of an improvised conversation or discussion, the researcher’s intention was to explore other ways of engaging with the idea of a conversation.

A recording from 2019, made by the researcher and a friend, was loaded into Ableton Live and transcribed using Ableton’s melody transcription software and a MIDI approximation was generated. The MIDI was assigned to two instrument racks that were programmed with cut-up speech sounds from Source A and Source D respectively. Drums were improvised along with the resulting samples.

The result of this experimentation was a very dense vocal texture that obscured any memorable thematic content. The original conversation that was transcribed exhibited a monotonous melodic phrasing as the dialogue was not particularly lively. This conversation was chosen because of its sentimental, archival nature and the researcher wished to develop a piece from an artifact of their lived experience.

The researcher still wished to explore the ways in which conversation or natural speech rhythms could be used to inform thematic content. However, the lack of development and imperceptibility of the first iteration of the piece called for a reworking of the structure and melodic content.

The melodic content of Source A was edited to create clearer phrases with breaks to differentiate melodic ideas. The cut-up speech sounds from the first version of the piece were used in the reworked version as accompaniment and worked with the re-recorded drums to build a foundation. A lead line was created from the same cut-up speech sounds but focusses on providing the listener with fragments of a relatively cohesive cut-up phrase (one that contains enough of the original speech to suggest linguistic meaning) during the first few sections of the piece (A-D) and is only revealed in-full once the piece reaches section E. The instruments used to produce the cut-up speech sounds for Source A and D take each utterance of the original cut-up and assign them to a note that can be triggered using MIDI controllers. This means that one could play the cut-up pieces in order and, depending on the rhythm used,

create a very close approximation of the original phrase spoken by the source. The composer enjoyed the idea that as the piece unfolds as the listener is presented parts of the original phrase that develop the context of the cut-up speech sounds used for the accompaniment. At 3m 44s the listener is presented with a section that contains only the lead voices playing the original phrases in their entirety, interacting as two people conversing, before devolving back into the texture of the first three minutes of the piece. However, the listener now has more context about the material used and can potentially interpret meaning based on how the fragments interact out of sequence. The piece was given more defined formal structure that is outlined by the shift in style of the drums from section to section.

6.11 Whatever It Takes

"Whatever It Takes", as the title suggests, is a track that focusses on resilience. The perspective of the track is that of a subject fantasising about what it might be like to have unwavering confidence and determination when faced with adversity. How might it feel to be met with obstacles and be fearless in pursuit of forward momentum? "Whatever It Takes" felt like a triumph for the researcher. It was created at a point when inspirational ideas were challenging to generate and reinvigorated the researcher's motivation and self-belief that was beginning to fade. It appears as the last track and represents the energetic and symbolic climax of the album.

Table 6. 11: "Whatever It Takes" Formal Structure

Form	Intro	A	B	C	B	C	D	C	A'	A''	Outro
Length (Bars)	17	10	31	25	10	8	36	18	10	16	8
Tempo	163 bpm										
Time Sig.	4/4										
Duration	4' 35''										
Instrumentation	Drums & Source D										

The first version this piece was a result of heavily processing the cut-up speech sounds of Source D. The sounds were programmed into an Ableton Live instrument rack that would trigger the sounds with varied levels of distortion based on the velocity with which one hit the sample pad. The higher the velocity, the more distortion affected the samples.

The initial version of the piece was created from an improvisation using the Ableton Live instrument and drums. The result was a heavily groove-based piece. The timbre of the cut-up speech sounds influenced the pseudo-punk drumming style of the accompaniment.

The initial piece lacked structure and variation and the intensity of the material quickly diminished. The improvisation was recorded without a click, the result of which was a "sloppy" sound that somewhat fitted with the aesthetic of the material but obscured the rhythmic ideas of both the voice and the drums.

The MIDI from the first version of the track was quantised to the smallest available note value (thirty-second notes). The researcher then reviewed the vocal material for thematic content that could act as the foundation for sections. Appropriate themes were arranged to create the overall form. Once in place, the thematic content was edited using Ableton's time-stretching software to create longer and shorter note values to use for contrasting material and as fills that could break the monotony of sections or bridge to thematic ideas. The drums were re-recorded and follow the vocal foundation directly for much of the track. Additional melodic material was added to the sections in an accompanying role (also Source D).

The track draws influence from the Nine Inch Nails album *With Teeth* (2005), specifically the track "Hand That Feeds". Trent Reznor's timbral aesthetic in "Hand That Feeds" also focusses on driving and repetitive rhythmic phrases, and relatively short melodic ideas. The resulting sense of the track is that of unrelenting momentum.

6 Conclusion

This dissertation concludes by discussing the extent to which the research was able to answer the primary research questions. For each question the researcher's methodological approach, the results of the qualitative and quantitative analysis, and the compositional practice are considered and reflected upon. The researcher also reflects on his creative process (the subjective artistic benefits and challenges of a compositional undertaking in the manner detailed) and some possibilities for further research.

Do speech and music share parameters that express emotive meaning in an analogous manner?

Speech and music are both sonic-based modes of communication that can be structurally interpreted in terms of the following parameters: pitch, rhythm, timbre, and dynamics. In both forms of communication, these four parameters act as facilitators of the medium. In speech, the semantic information conveyed by combining phonemes into morphemes and organised by syntax, is only able to be expressively conveyed in combination with these parameters. In music, a composer or performer's knowledge and their technical ability to manipulate these parameters is the manner through which they express, regardless of their aesthetic intentions. Therefore, as receivers of speech and musical communication, we experience these parameters, often in combination with other contextual elements and decode the nature of their use based on our prior experience.

Aspects of this research question were interrogated by divorcing semantic meaning from speech and implementing an adapted audio version of the literary cut-up technique to create brief audio clips of composed cut-up speech sounds. These were subsequently presented to participants to gather qualitative data that would develop the researcher's understanding of the expressive nature of cut-up speech sounds. The subsequent analysis of the collected data shows that participants were able to interpret meaning, but the meaning was highly subjective in nature, varying widely across the sample, and only in select instances similar to the expressive content of the original phrases. Therefore, the researcher's chosen methodological approach does not claim that the way one receives emotive meaning through speech and music is analogous in nature, only that both mediums as a result of their sonic and temporal nature, manipulate pitch, rhythm, timbre, and dynamics in service of conveying meaning.

How does the act of composing with cut-up speech sounds change the emotive meaning for the receiver?

When presented with examples, the responses showed that for certain examples some participants were able to interpret the emotive delivery of the original phrase, while for other examples the emotive delivery was altered by the compositional process. The variation and lack of significant correlation across the participant responses illustrates that the results do not determine precisely how the emotive meaning changes through the composition process, only that the contextual change in the relationship between cut-up speech sounds as a result of the process, alters the emotive meaning interpreted by a listener. This notion inspired the researcher to experiment with and manipulate the cut-up speech sounds in the musical works with the knowledge that a listener would be able to interpret emotive expression, but not in a way that would be predictable or controllable.

Can a composer extract and harness the speech sounds which allow one to signify emotive meaning through spoken language?

The initial desire of this research was to produce a taxonomy of cut-up speech sounds that could be utilised and give a listener musical works that expressed meaning as a speech surrogate. A composer could use the taxonomy to reproduce the sensation of emotive speech. In pursuit of this initial goal, the research did not lead to usable results given the first proposed taxonomic approach. For this to be possible, the amount of data required would have to go significantly beyond what this research was able to capture and the nuance within the data would need to detail the subjective nature of each participant's response to each individual speech sound in isolation and in various contrasting musical contexts.

However, in practice the researcher was able to create emotively expressive music using cut-up speech sounds without utilising a rigid taxonomy. By divorcing speech from its semantic meaning, the researcher was able utilise the ambiguity of musical expression to create accessible (insofar as the timbre of the source material and structural elements governing the arrangement of musical ideas are those that a listener would likely be able to identify) and emotive musical works.

Directions for further research

It is the researcher's belief that further exploration into the relationship between how speech and music convey emotive meaning could yield insightful research detailing the psychological experience of the two modes of communication. For this to be effective and generate significant data, responses will need to be gathered from participants from various cultural contexts and language demographics in relation to systematically and meticulously generated and isolated speech utterances. The scope of such research would take multiple iterations of surveys and interviews.

The creative process of composing with cut-up speech sounds proved to be rewarding due to the creative limitations imposed. Choosing to compose with limited raw materials was artistically challenging; the process encouraged the researcher to examine the characteristics of the cut-up speech sounds in a detailed and enlightening way. Speech sounds contain extraordinary timbral detail and the manipulation of their inherent parameters resulted in rich creative material. The vocal, and thus, the expressive quality of the raw material was able to withstand rigorous editing and manipulation which allows the music to retain a human quality, one that the researcher took care to preserve throughout the process. As a musician that is privileged enough to have access to the digital music technology of modern music production, limiting oneself in terms of instrumentation forces a nuanced exploration of the potential of sound and is a worthwhile undertaking for any musician wishing to deepen their understanding of the possibilities of electronic production.

This research has illuminated that the methodological approach produced results that are inter-subjective and inter-contextual in nature. The emotive expression of the cut-up speech sounds was not similarly received across the sample set. However, creating a musical taxonomy of cut-up speech sounds remains a potential field for artistic enquiry, but one that would require a different methodological approach than the approach of this dissertation. The researcher has shown that the methodological approach detailed in this research did not yield the expected musical taxonomy, but it is possible that a more extended project may yet be able to do so.

References

Adorno, T. W. 2002. *Essays on Music: Theodore W. Adorno*. Berkeley, California: University of California Press.

Barrett, E. 2007. "Introduction". In *Practice as Research: Approaches to Creative Arts Enquiry*, ed. Estelle Barrett and Barbara Bolt. London: I.B. Tauris, 1-13.

BBC. 2015. "What is the Cut-Up Method?". https://www.bbc.com/news/magazine-33254672, accessed 22 February 2024.

Blythe, S. 1994. "Karl Pearson and the Correlation Curve", *Internal Statistical Review*, 62(3), 393-403.

Bock, Z. 2014. "Approaches to Communication". In *Language, Society and Communication: An Introduction*, ed. Bock, Z., and G. Mheta. Pretoria: Van Schaik Publishers, 35-54.

Boynton, P. M., and T. Greenhalgh. 2004. "Hands-on guide to Questionnaire Research: Selecting, Designing, and Developing your Questionnaire", *BMJ: British Medical Journal*, 328(7451), 1312-1315.

Braae, N. 2014. (Review) "The Accessibility of Music: Participation, Reception, and Contact", *Popular Music and Society*, 395-398. http://dx.doi.org/10.1080/03007766.2014.943946, accessed 22 February 2024.

Braun, V., and V. Clarke. 2013. *Successful Qualitative Research: A Practical Guide for Beginners*. India: SAGE Publications.

Burroughs, W. S. 2014 [1962]. *The Ticket That Exploded*. London: Penguin Classics.

Chrysagis, E. 2014. (Review) "The Accessibility of Music: Participation, Reception, and Contact by Jochen Eisentraut", *Popular Music*, 33(2), 367-370.

Crispin, D. 2015. “Artistic Research and Music Scholarship: Musings and Models from a Continental European Perspective”. In *Artistic Practice as Research in Music*, ed. M. Doğantan-Dack . New York: Routledge.

Davis, M. 2012. “The Music of Reason in Rousseau’s ‘Essay on the Origin of Languages’”, *The Review of Politics*, 74(3), 389-402.

Dediu, D. 2012. “Modernity and Accessibility. A Possible Paradigm for Musical Composition”, *Lucrări de* Muzicologie, 2, 48-67.

Dhakal, K. 2022. “NVivo”, *Journal of the Medical Library Association*, 110(2), 270-272.

Dilley, L. C. 2009. (Review) “Music, Language, and the Brain by Aniruddh D. Patel”, *Phonology*, 26(3), 535-540.

Eisentraut, J. 2013. *The Accessibility of Music: Participation, Reception, and Contact*. United Kingdom: Cambridge University Press.

Emmerson, S. 1986. “The Relation of Language to Materials”. In *Language of Electronic Music*. London: Palgrave MacMillan.

Feld, S., and A. A. Fox. 1994. “Music and Language”, *Annual Review of Anthropology*, 23, 25-53.

Frayling, C. 1993-1994. “Research in Art and Design”, *Royal College of Art Research Papers* 1(1), 1-5.

Ganti, A. 2023. “Central Limit Theorem (CLT): Definition and Key Characteristics”. https://www.investopedia.com/terms/c/central_limit_theorem.asp, accessed 24 January 2024.

Gracyk, T. 2003. "Sound and Vision: Color in Visual Art and Popular Music". In *Pop Sounds: Klangtexturen in der Pop- und Rockmusik. Basics - Stories – Tracks*, eds. Phleps, T., and R. von Appen, 49-64. https://doi.org/10.1515/9783839401507-005, accessed 24 January 2024.

Hannula, M., J. Suoranta and T. Vadén. 2014. *Artistic Research Methodology: Narrative, Power and the Public*. Austria: Peter Lang.

Jamieson, S. 2023. "Likert scale". *Encyclopaedia Britannica*. https://www.britannica.com/topic/Likert-Scale, accessed 24 January 2024.

Kader, G. D., and C. A. Franklin. 2008. "The Evolution of Pearson's Correlation Coefficient", *The Mathematics Teacher*, 102(4), 292-299.

Kafle, S. C. 2019. "Correlation and regression analysis using SPSS", *Management, Technology & Social Sciences*, 126-132. https://www.researchgate.net/publication/343282545_Correlation_and_Regression_Analysis_Using_SPSS?enrichId=rgreq-6596833909670e53f8571eb8b5b23498-XXX&enrichSource=Y292ZXJQYWdlOzM0MzI4MjU0NTtBUzo5MTg1ODg1NDYzNjMzOTJAMTU5NjAxOTk4NDQ4Ng%3D%3D&el=1_x_2&_esc=publicationCoverPdf, accessed 25 January 2024.

Kendall, M. G., 1938. "A New Measure of Rank Correlation", *Biometrika*, 30(1/2), 81-93.

Kennett, C. 2008. "A Tribe Called Chris: Pop Music Analysis as Idioethnomusicology", *Open Space Magazine*, 10, 8-19.

Koven, M. 2014. "Interviewing: Practice, Ideology, Genre, and Intertextuality", *Annual Review of Anthropology*, 2014(43), 499-520.

Langer, S. K. 1953. *Feeling and Form: A Theory of Art*. New York: Charles Scribner's Sons.

Langer, S. K. 1964. "Abstraction in Art", *The Journal of Aesthetics and Art Criticism*, 22(4), 379-392.

Langer, S. K. 1980 [1942]. *Philosophy In a New Key: A Study in the Symbolism of Reason, Rite, and Art*. Cambridge, Massachusetts: Harvard University Press.

Monelle, R. 1992. *Linguistics and Semiotics in Music*. Chur: Harwood Academic.

Murphy, T. S. 2002. "The Nova Trilogy". *The Literary Encyclopedia*. https://www.litencyc.com/php/sworks.php?rec=true&UID=10822 , accessed 25 January 2024.

Negus, K. 2012. “Narrative, Interpretation, and the Popular Song”, *The Musical Quarterly*, 95(2/3), 369-395.

Nelson, R. 2013. *Practice as Research in the Arts: Principles, Protocols, Pedagogies, Resistances*. United Kingdom: Palgrave Macmillan.

Odier, D. 1970: *The Job: Interviews with William S. Burroughs*. New York: Grove Press.

Qu, S. Q., and J. Dumay. 2011. “The Qualitative Research Interview”, *Qualitative Research in Accounting & Management*, 8, 238-264.

Reece, S. 1977. “Forms of Feeling: The Aesthetic Theory of Susanne K. Langer”, *Music Educators Journal*, 63(8), 44-49.

Rushdie, S. 2006 [1981]. *Midnight’s Children*. London: Vintage.

Ryan, C. T. S. 2024. “Cut ups”. https://www.briongysin.com/cut-ups/, accessed 25 January 2024.

Simon, E., and T. Kunkeyani. 2014. “Phonetics: the study of human speech sounds”. In *Language, Society and Communication: An Introduction*, eds. Bock, Z., and G. Mheta. Pretoria: Van Schaik Publishers, 83-118.

Smalley, D. 1997. “Spectromorphology: Explaining Sound-shapes”, *Organised Sound*, 2(2), 107-126.

Spearman, C. 1904. “The Proof and Measurement of Association between Two Things”, *The American Journal of Psychology*, 15(1), 72-101.

Young, J. 2015. “Imaginary Workspaces: Creative Practice and Research through Electroacoustic Composition”. In *Artistic Practice as Research in Music*, ed. Mine Doğantan-Dack. New York: Routledge.

Music

Flying Lotus. 2019. “Fire is Coming”. *Flamagra*. Warp Records.

Heernt. 2006. “Locked in a Basement”. *Locked in a Basement*. Razdaz Recordz.

Mark Guiliana Jazz Quartet. 2017. “Inter-Are”. *Jersey*. Motema.

Nerve. 2011. “Tetrastigm”. *The Distance Between Zero and One*. BNS Sessions.

Newman, R. 1995. “You’ve Got a Friend in Me”. *Toy Story (Motion Picture Soundtrack)*. Walt Disney Records/Pixar.

Nine Inch Nails. 1994. “Hurt”. *The Downward Spiral*. Nothing/Interscope Records.

Nine Inch Nails. 2005. “Hand That Feeds”. *With Teeth*. Interscope Records.

Radiohead. 2000. “Everything In Its Right Place”. *Kid A*. XL Recording Ltd.

Smith, N. 2018. *Pocket Change*. Ropeadope LLC.

Tobin, A. 2002. “Verbal (feat. MC Decimal R.)”. *Out From Out Where*. Ninja Tune.

Tobin, A. 2011. *ISAM*. Ninja Tune.

Vic Firth. 2021. “Performance Spotlight: Jojo Mayer & Jack Quartet – Different Zones by Don Li”. https://youtu.be/MhV9h9K8yHo, accessed 23 February 2024.

Zimmer, H., and B. Wallfisch. 2017. *Blade Runner: 2049 (Original Motion Picture Soundtrack)*. Alcon Sleeping Giant (ASG) Records.

Appendices

A1: Example Rationales

A1.1: Example 1

The original audio was split into 16 pieces of which 9 were removed. The cut-ups were chosen to try and accentuate the sombre tone in the original audio. The tone of the voice is low-mid heavy, rich, and full sounding. The researcher chose to accentuate the vowel sounds of the phrase, removing the consonant sounds at the beginning of each word. As a result, the edit contains very little of the percussive initial transient onset characteristic of consonant sounds which, by design, means that the example has a smooth quality. The tape delay used on this example helps to create the illusion of depth and space for the sound. The tape delay is affecting the range between 30-870 Hz which emphasises the low-mid heavy character of the voice. The goal of this example was to create a dense and rich sounding voice.

A1.2: Example 2

The original audio was split into 19 parts and 10 were removed. The researcher placed two sombre examples together in the order to see how the listener's reaction would change between a male and female source-sound. There is more sibilance in this example than the previous. The example is looped to further juxtapose the previous example. A high shelf equalization (EQ) is used to make the sound brighter and more present and the mid to high-mid range is also sent through the tape delay to emphasise this range of the vocal. The delay also creates a sense of movement to avoid sounding static or lifeless. The feedback on the tape delay is used to create a high-end cloud that is present throughout playback. Modulation augments the voice and, perhaps, creates a bit of an unnatural feeling. The loop, coupled with the low end naturally present in the male voice, creates an implied regular, metric feel.

A1.3: Example 3

The original audio was split into 9 parts and 4 were removed. The tone of the original audio was that of anger. The example is brief, direct, dry, and recorded in close proximity to the microphone to convey a visceral experience of the anger in the voice. To convey this emotion in the edited version of the recording, a phrase recorded close to the microphone was used. The immediacy of the sound accentuates the emotive quality of the audio. The rhythm of this example has been left unchanged to reflect the natural rhythm of the transient onset in the

speech. In theory, the original rhythm was retained to impart as much of the original emotion as possible. The rhythm is rapid and syncopated which adds to the uneasy tone of the voice. There are a mix of consonant and vowel sounds from the original phrase in the cut-up which was intended to give the listener a slightly clearer picture of the original audio. The researcher employed a minimalist approach to the processing of the example. Very little, apart from the cut-up technique, has been used to alter the voice.

A1.4: Example 4

The original audio was split into 12 parts and 6 were removed. The researcher opted for a "clear but soft" attack in this example, aiming for the voice to sound somewhere closer to rolling or flowing than normal speech. The example is reversed to limit the linguistic understanding and place emphasis on the tone instead. The example contains mostly vowel sounds, but is punctuated by consonants, most noticeably the "t" and "s". These sounds imply a sense of rhythm because of their percussive nature and create the impression that there is a phrase with multiple words in the example. The delay is used to give movement and depth to the sound. The low to low-mid range that is sent through the tape delay adds an artificial pulse or vibration to the sound that creates rhythms rather than regular beats. The aim of the effects placed on the audio was to create a dynamic sounding example that would contrast the neutral emotive delivery of the original audio.

A1.5: Example 5

The original example was split into 29 parts and 16 were removed. This example utilises the percussive sibilant and plosive sounds present in the original audio so that the utterances are short and sharp. The more melodic quality of the longer vowel sounds was removed to create a clearly defined rhythm. The percussive and rhythmic characteristic of the example was created to heighten the emotion of the original phrase. The rationale was that the harsh nature of the rhythm and the tone of the sibilant and plosive sounds would create an uncomfortably intimate feeling in the listener. The example is heavily compressed, and the high-end range of the voice is put through a tape delay to further accentuate the harsh nature of the sound. When creating the example, the researcher imagined a person vibrating from their anger.

A1.6: Example 6

The original audio was split into 12 parts and 7 were removed. The original phrase was chosen because it exhibits anger but is delivered in a whispered tone. The aim for this example was to

accentuate the sounds of the breath present in the original delivery and to obscure the linguistic information. To ensure that the original phrase was as unrecognisable as possible, the audio was reversed. As a result, this example is tone focused. A bit-crusher is used to create an uncomfortable and close sound that feels extremely present and rough to increase the anger and aggression in the voice. Tape delay that focuses on the low-end (60-460 Hz) is used to accentuate the vibrant low-end that is present in the whispering voice. The delay also has the effect of making the sound feel less real; more imaginary, ghostly, and unsettling.

A1.7: Example 7

The original audio was split into 23 parts and 14 were removed. The researcher tried to create a musical melody from the original audio by accentuating the contour from low to high found in the speech. The rhythm that remains implies a meter to enhance the musical nature of the phrase. Many of the percussive sounding consonant sounds were removed, leaving behind brief vowel and non-percussive consonant sounds to mimic the character of staccato wind or brass instruments. The emotion in the original audio is excitement and joy, so the researcher used a flanger audio effect to create a sense of bouncing in the sound. The rhythm from the original recording is unchanged, which helps to suggest an energetic and lively mood.

A1.8: Example 8

The original audio was split into 15 parts and 9 were removed. The aim of this example was to subvert the anger that is present in the original. The researcher's intention was to explore the degree to which the sound of the voice could be altered to change that emotional characteristic of the phrase. The original recording had intervallic movement, meaning that the phrase did not feel static or monotonous. The example accentuates this intervallic movement by creating a more rhythmic and staccato phrase. This, in combination with the looped nature of the example, implies a more musical structure than that of the original recording. The original audio has been pitch shifted down by the process of time-stretching with the intention to elongate the sound and make it sound more artificial. The delay affects the low-mid to mid-range of the sound but has very little feedback and is not very wet in the mix. The delay is used to imply size, both of source and space. The rhythm is unchanged from the original recording so that cut-ups sound at the same rate as the original speech. The percussive consonant sounds of the voice are not emphasised to see whether this will affect the listener's ability to perceive the emotion of the original clip. The flanger is the most present effect on the sound and helps

to accentuate the tone-colour rather than the language of the sound. The effect also works to emphasize the dynamic and rhythmic feeling of the phrase.

A1.9: Example 9

The original audio was split into 12 parts and 6 were removed. The overall sound of this example focuses on the powerful, percussive “k” consonant sound in the original recording. The low-end of the original audio has been removed to further accentuate the powerful, crisp, consonant sound. The rhythm is pronounced and clear because of the short nature of the clips. The delay is five seconds long, which means that by the time the audio is repeated in the test one could have heard a faint of echo of the cut-up, much softer, and more palatable than the primary iteration. Low feedback and low mix levels ensure that the delayed repetition is far less present than the primary iteration.

A1.10: Example 10

The original audio was split into 12 parts, and none were removed. The order of the split parts was rearranged. There is minimal processing, apart from the cut-up technique. The pitches of the vowels are the focal point of the example. This example exhibits intervallic jumps which have been maintained to retain the happy emotive character of the original recording. There are repeated clips in the cut-up which have been included as constants in the phrase. There are three distinct phrases, established by two distinct pauses or rests in the sound. This implies a feeling of call and response in the phrase.

A1.11: Example 11

The original audio was split into 7 parts and 3 were removed. The researcher wanted to diminish the happy emotive delivery of the original audio for this example. The consonant sounds from the beginning of the words in the original recording were removed, as well as the low and extreme high-end frequency range. The character of the voice feels restricted as a result. The example feels icy and cold, created using two tape delays set to a rapid delay time with moderate to high feedback, which have been set almost completely wet in the mix (meaning that more of the audio effect is present). This creates a shrill and energetic sound. The second delay is set to act as an echo, which gives the feeling of space and size.

A1.12: Example 12

The original audio was split into 8 parts and 4 were removed. This example is looped to see if it influences whether it is easier for a participant to hear a phrase. The low-mid range is accentuated by both the EQ and tape delay. There is a very present "heartbeat" pulse created by the delay (high feedback and relatively fast delay time). It is very wet in the mix, so the audio is very clearly altered from the original. The proximity of the whisper and the richness of the sound at a close distance to the microphone greatly influenced how the sound is cut-up. The aim was to mimic the act of close-micing a drum-set. Every sound is extremely intimate (too intimate potentially). The original cut-up rhythm (before delay) is very regular which helps create to a pulse which the delay can augment and accentuate. The percussive consonant sounds ("c", "t", "s") were isolated and retained from the original recording, which further adds to the percussive and highly rhythmic feeling of the phrase.

A1.13: Example 13

The original audio was split into 30 parts and 10 were removed. Apart from the sibilant "s" consonant in the original phrase, much of the percussive consonant sounds have been removed. The example has been reversed to investigate how it would affect the way a participant is able to perceive the original emotion of the speech. The rhythm of the utterances remains unchanged from the original recording. It feels rapid, and the tone is monotonous. Little processing was used to keep the audio as dry as possible to see what effect the reverse would have as the only effect applied.

A1.14: Example 14

The original audio was split into 54 parts and 29 were removed. The aim of using the original recording was to derive a series of phrases that when cut-up would create a lengthy example. The beginnings of each word from the original recording were used to create this example: both consonant and vowel sounds. Many of the words begin with consonant sounds implying a clear rhythmic phrase. The example makes use of two cut-up phrases to imply an antecedent and consequent phrase structure (call and response). The original audio exhibits a melodic contour that descends in pitch from beginning to end. Tape delay was used with a 976ms delay time (roughly one second) to create a denser rhythmic phrase as the example unfolded. Due to the word fragments used in this example, the linguistic information from the original recording can be partially understood. Thus, the delay also creates the sense that the original meaning is slowly being removed as the rhythmic texture becomes more chaotic.

A1.15: Example 15

The original audio was split into 14 parts and 8 were removed. This example places emphasis on, and accentuates, the timbral quality of the original audio. The original recording was delivered in a whispered tone with the source extremely close to the microphone. As a result, the microphone was able to capture, not only the breath of the whisper in detail, but also the sounds one's mouth makes when speaking. The example makes use of these sounds to convey the intimacy (regarding space) of the voice. The audio was slightly time-stretched to elongate the breathy quality of the delivery and as a result, has been lowered in pitch from the original audio. An EQ was placed on the audio, boosting the frequencies around 4.1 kHz using a notch filter, to further accentuate the "hiss" of the breath.

A1.16: Example 16

The original audio was split into 15 parts and 8 were removed. The vowel sounds of the original recording were used to create this example. To accentuate the happy emotive quality of the delivery in the original recording the amount of percussive consonant sounds present in the example were limited. The delivery of the phrase was spoken in a "sing-song" fashion, which subsequently imparted a satisfying melodic contour onto the voice. The phrase rises and falls in pitch which, coupled with the lack of percussive consonant sounds in the example, aim to create a phrase that feels musical while accentuating the contour and emotive expression in the original recording.

A1.17: Example 17

The original audio was split into 11 parts and 7 were removed. It is delivered in a happy and encouraging tone. The aim of this example was to diminish or subvert the expression of the original. The vowel sounds of the original were used but appear in very brief utterances. The example is looped to give an unrelenting and annoying quality. Many of the low frequencies have been filtered out to remove warmth from the voice. A tape delay of 233ms was placed on the audio to give a rhythmic and repetitive feeling, enhancing the unrelenting and annoying quality established by the loop.

A1.18: Example 18

The original audio was split into 31 parts and 17 were removed. The beginning consonant sounds of the words from the original recording were used to better convey the articulation of the speech. The delivery of the original recording was shouting in an angry tone. By retaining

the beginnings of the utterances, the example exhibits a clear rhythmic quality and the anger in the delivery is maintained. A high-pass shelf was placed in the audio to completely remove all frequencies below 500 Hz and with them the richness present on the low frequency range of the original recording. The result is a voice that appears to have a cold, lifeless, harsh, and percussive tone.

A1.19: Example 19

The original audio was split into 13 parts and 6 were removed. The objective with this example was to create a phrase that contrasted the original recording. The audio is heavily distorted to remove any tone colour or variation from the voice. The example has been reversed to further obscure the linguistic information present and suggest an unnatural delivery. The rhythm of the example is noticeable, but the quality of the tone is not percussive. The fact that the audio has been reversed means that the strong consonant sounds that start each utterance have been shifted to the end of each sound. The result is a phrase that is slow relative to its length.

A1.20: Example 20

The original audio was split into 9 parts and 5 were removed. The aim of this example was to accentuate the happy and excited emotive delivery of the original recording. The consonants that begin each word were retained to emphasize the rhythm and maintain the energy of the original. The phrase is brief, and the rhythm is relatively fast, implying excitement or energy. A tape delay of 104ms was placed on the audio to heighten the feeling of movement in the delivery. There are large intervallic jumps. A flanger was used to impart the feeling that the voice is bouncing or vibrating with each utterance.

A1.21: Example 21

The original audio was split into 13 parts and 7 were removed. The example is looped to investigate whether it is easier for a participant to hear a word or phrase relative to other examples. The beginnings of the words from the original recording have been removed, leaving behind mostly vowel utterances. The rationale for doing so was to prevent the example from being overtly percussive due to the strong “k” consonant sound from the original audio. The delivery of the original recording was sad, and the researcher edited and processed the audio in a manner that would augment that emotion rather than diminish it.

A1.22: Example 22

The original audio was split into 41 parts that comprise doubled clips. The parts were reordered.

A1.23: Example 23

The original audio was split into 54 parts and 29 were removed. The parts were then time-stretched to alter the pitch and reversed. The aim of this example was to augment the pitch of the original recording in an unnatural manner. The example exhibits very low- and very high-pitched utterances. The low-pitched sounds are due to the audio being time-stretched so that it became slower, and the high-pitched utterances are a result of tape delay accentuating the frequencies between 1 kHz and 20 kHz. The mix of the tape delay was set to 79% of the original audio, meaning that the affected high-pitched sounds would be perceived with the same intensity as the rest of the audio.

A1.24: Example 24

The original audio was split into 15 parts and 8 were removed. Utterances were chosen that began with vowel sounds and ended with consonant sounds. The delivery of the original recording was neutral, and the researcher worked to maintain that quality in the example. The rhythm is noticeable but appears neither fast nor slow. The voice was filtered with a high-pass EQ shelf that removes all frequencies below 300 Hz so that the tone of the voice would appear cold and lacking in colour and richness. A flanger was placed on the audio to give the impression that the voice is robotic or talking over the phone.

A1.25: Example 25

The original audio was split into 17 parts and 9 were removed. The aim of this example was to use the inherent tone of the original recording to diminish the angry nature of the delivery. The original recording displayed prominent sustained vowel sounds that were used to create the example. The utterances are not percussive but flow smoothly from one to the next. The example has a dynamic melodic contour that creates the feeling that one is listening to a lively melody. The looped nature of the example also aids in imparting a melodic and musical structure onto the example.

A1.26: Example 26

The original audio was split into 29 parts and 10 were removed. The example was created to diminish the happy emotive delivery of the original, but not by creating a phrase that felt harsh

or percussive. Rather, the audio is affected in such a way as to obscure the voice completely; to create an example that would not be recognisable as voice. The utterances are brief, so that little to no linguistic information could be imparted and arranged in an even pulse rather than following the rhythm of the original recording. There is a tape delay of 500ms was used to create a metered feeling. A second tape delay accentuates the low frequencies (also with a delay time of 500ms) to add a pulsing feel to the audio. As the example plays the delay also creates a denser sounding texture further obscuring the utterances . The audio was time-stretched slower and subsequently pitched down. The tape delay, much like example 23, creates contrast in the form of extreme high-pitched sounds that counter the low rumble of the audio. The impression is that of multiple sound sources rather than just one voice.

A2.: Supplementary Results

Table A2.1: Examples with Anthropological Descriptions (Participant Survey)

	A : Age	B : Gender	C : Languages	D : Locations	E : Occupation	F : People
1 : Ex 1	0	4	3	0	0	1
2 : Ex 2	1	3	3	0	0	1
3 : Ex 3	1	4	4	0	0	2
4 : Ex 4	0	4	2	0	0	1
5 : Ex 5	0	8	0	0	0	1
6 : Ex 6	1	5	3	0	0	2
7 : Ex 7	2	8	6	0	1	0
8 : Ex 8	0	3	1	0	0	0
9 : Ex 9	0	1	5	1	1	2
10 : Ex 10	1	7	16	2	3	7
11 : Ex 11	1	12	0	0	1	0
12 : Ex 12	0	5	0	0	0	0
13 : Ex 13	1	11	5	0	1	0
14 : Ex 14	0	3	1	0	0	0
15 : Ex 15	2	3	0	0	0	0
16 : Ex 16	2	8	1	0	0	1
17 : Ex 17	1	5	1	0	3	0
18 : Ex 18	0	3	2	0	1	2
19 : Ex 19	0	3	0	0	0	0
20 : Ex 20	8	5	0	0	0	0
21 : Ex 21	0	4	1	0	0	0
22 : Ex 22	0	5	4	0	2	1
23 : Ex 23	0	2	0	0	0	0
24 : Ex 24	3	7	1	0	1	2
25 : Ex 25	3	10	0	0	0	1
26 : Ex 26	1	0	0	0	1	1

Table A2.2: Examples with Emotion (Participant Survey)

	A : Anger	B : Anticipation	C : Fear	D : Joy	E : Loathing	F : Sadness	G : Surprise
1 : Ex 1	0	3	7	3	0	1	1
2 : Ex 2	3	1	7	3	0	0	0
3 : Ex 3	4	2	0	3	2	0	2
4 : Ex 4	1	2	13	2	2	2	0
5 : Ex 5	4	1	7	8	2	0	0
6 : Ex 6	0	3	33	0	0	1	0
7 : Ex 7	6	0	4	11	1	0	1
8 : Ex 8	1	2	17	3	1	0	0
9 : Ex 9	0	1	2	3	1	2	0
10 : Ex 10	1	0	2	6	1	0	0
11 : Ex 11	1	0	19	2	1	1	2
12 : Ex 12	0	1	9	8	0	0	0
13 : Ex 13	1	2	5	4	0	2	0
14 : Ex 14	5	0	2	2	8	0	0
15 : Ex 15	0	0	23	2	0	1	1
16 : Ex 16	1	0	0	11	1	1	0
17 : Ex 17	5	0	4	4	2	0	0
18 : Ex 18	5	0	4	2	1	4	0
19 : Ex 19	5	1	17	0	0	1	0
20 : Ex 20	0	0	4	39	0	0	0
21 : Ex 21	1	4	4	3	1	0	0
22 : Ex 22	1	0	1	1	0	0	2
23 : Ex 23	1	0	12	2	2	0	1
24 : Ex 24	1	0	1	3	0	3	1
25 : Ex 25	1	1	2	6	0	0	0
26 : Ex 26	0	2	12	4	1	0	0

Table A2.3: Examples with Sonic Descriptions (Participant Survey)

	A : Cut up	B : Movement	C : Speed	D : Texture	E : Tone	F : Unclear	G : Volume
1 : Ex 1	6	5	0	0	2	12	0
2 : Ex 2	1	4	1	4	0	3	2
3 : Ex 3	1	1	13	3	0	4	0
4 : Ex 4	0	4	1	0	2	7	1
5 : Ex 5	20	4	9	3	1	2	0
6 : Ex 6	0	0	0	1	1	3	1
7 : Ex 7	12	6	2	1	1	3	0
8 : Ex 8	1	2	1	1	6	1	0
9 : Ex 9	8	11	4	2	0	1	8
10 : Ex 10	3	0	0	0	1	1	0
11 : Ex 11	1	2	0	2	3	4	4
12 : Ex 12	1	9	3	3	1	0	0
13 : Ex 13	1	3	2	1	1	4	1
14 : Ex 14	5	2	3	13	0	6	0
15 : Ex 15	0	1	0	1	2	1	0
16 : Ex 16	4	4	0	0	0	0	1
17 : Ex 17	8	3	4	1	1	1	0
18 : Ex 18	17	2	4	2	2	2	0
19 : Ex 19	1	5	1	5	6	0	1
20 : Ex 20	1	2	0	1	1	0	0
21 : Ex 21	2	5	4	3	4	1	1
22 : Ex 22	8	2	7	4	0	11	0
23 : Ex 23	1	0	4	11	1	11	0
24 : Ex 24	8	1	2	1	0	4	0
25 : Ex 25	3	3	1	2	0	2	1
26 : Ex 26	0	5	2	2	4	7	0

Table A2.4: Expert Examples with Anthropology Descriptions (Expert Interviews)

	A : Age	B : Gender	C : Location & Language	D : Occupation	E : People
1 : Ex 1	0	1	5	0	0
2 : Ex 2	0	1	0	0	0
3 : Ex 3	0	0	1	0	2
4 : Ex 4	1	2	0	0	1
5 : Ex 5	1	1	1	0	0
6 : Ex 6	0	0	2	0	1
7 : Ex 7	0	4	1	0	2
8 : Ex 8	0	2	0	1	1
9 : Ex 9	1	2	7	1	1
10 : Ex 10	1	2	9	0	2
11 : Ex 11	0	4	0	0	0
12 : Ex 12	0	2	0	0	0
13 : Ex 13	0	3	5	0	1
14 : Ex 14	1	1	4	1	2
15 : Ex 15	1	3	0	0	0
16 : Ex 16	0	1	2	0	2
17 : Ex 17	0	2	0	1	0
18 : Ex 18	0	1	1	5	0
19 : Ex 19	0	1	0	0	0
20 : Ex 20	5	4	1	1	0
21 : Ex 21	0	1	0	1	0
22 : Ex 22	0	3	3	0	0
23 : Ex 23	0	0	1	0	0
24 : Ex 24	1	2	1	0	1
25 : Ex 25	1	3	0	0	1
26 : Ex 26	0	1	0	0	0

Table A2.5: Expert Examples with Emotion (Expert Interviews)

	A : Anger	B : Anticipation	C : Disgust	D : Fear	E : Joy	F : Love	G : Neutral	H : Sadness	I : Surprise
1 : Ex 1	1	0	1	2	1	2	0	0	0
2 : Ex 2	1	1	0	2	1	0	0	0	1
3 : Ex 3	9	1	0	1	2	0	0	0	0
4 : Ex 4	0	0	0	0	1	0	0	1	0
5 : Ex 5	8	0	0	5	3	0	0	0	0
6 : Ex 6	0	1	0	9	1	0	0	0	0
7 : Ex 7	3	0	0	1	8	0	0	0	0
8 : Ex 8	1	0	2	1	2	0	0	0	0
9 : Ex 9	0	0	0	0	0	0	0	0	0
10 : Ex 10	0	0	0	0	2	0	1	2	0
11 : Ex 11	0	0	0	6	4	0	0	3	0
12 : Ex 12	3	1	0	1	0	0	0	1	0
13 : Ex 13	0	0	0	0	4	0	0	1	0
14 : Ex 14	1	0	1	0	1	0	0	0	0
15 : Ex 15	0	0	0	16	1	0	0	0	0
16 : Ex 16	1	0	1	0	4	0	0	0	1
17 : Ex 17	6	0	1	1	3	0	0	0	0
18 : Ex 18	2	0	0	4	1	0	0	0	0
19 : Ex 19	0	0	0	14	0	0	0	0	0
20 : Ex 20	0	0	0	1	20	0	0	0	0
21 : Ex 21	0	1	1	0	0	0	0	0	0
22 : Ex 22	1	0	0	0	5	0	0	0	1
23 : Ex 23	0	0	2	4	1	0	2	0	0
24 : Ex 24	1	0	1	1	2	0	1	0	0
25 : Ex 25	3	0	0	0	1	0	0	0	0
26 : Ex 26	1	0	0	5	0	0	0	0	1

Table A2.6: Expert Examples with Speech (Expert Interviews)

	A : Mouth Sounds	B : Tone of Voice
1 : Ex 1	0	5
2 : Ex 2	2	5
3 : Ex 3	0	4
4 : Ex 4	0	5
5 : Ex 5	0	2
6 : Ex 6	0	6
7 : Ex 7	1	2
8 : Ex 8	1	1
9 : Ex 9	1	10
10 : Ex 10	0	9
11 : Ex 11	0	2
12 : Ex 12	1	11
13 : Ex 13	0	10
14 : Ex 14	0	4
15 : Ex 15	0	6
16 : Ex 16	0	4
17 : Ex 17	1	3
18 : Ex 18	1	6
19 : Ex 19	0	2
20 : Ex 20	1	2
21 : Ex 21	1	6
22 : Ex 22	1	7
23 : Ex 23	0	3
24 : Ex 24	0	11
25 : Ex 25	1	5
26 : Ex 26	1	1

Table A2.7: Expert Examples with Sonic Descriptions (Expert Interviews)

	A : Cut Up	B : Movement	C : Speed	D : Texture	E : Tone	F : Unclear
1 : Ex 1	0	0	0	0	0	0
2 : Ex 2	4	0	0	2	0	1
3 : Ex 3	6	4	1	0	0	0
4 : Ex 4	0	0	1	0	1	7
5 : Ex 5	1	2	2	1	1	1
6 : Ex 6	0	0	1	0	1	0
7 : Ex 7	3	3	0	0	0	0
8 : Ex 8	2	1	1	0	3	0
9 : Ex 9	0	2	0	1	2	0
10 : Ex 10	3	1	0	0	2	0
11 : Ex 11	0	0	0	1	3	2
12 : Ex 12	0	2	0	0	1	0
13 : Ex 13	0	0	0	0	0	2
14 : Ex 14	3	1	2	0	0	1
15 : Ex 15	0	0	1	0	5	0
16 : Ex 16	2	0	1	0	0	1
17 : Ex 17	3	1	1	0	1	0
18 : Ex 18	4	1	0	1	0	0
19 : Ex 19	0	1	1	3	2	3
20 : Ex 20	0	4	0	0	2	1
21 : Ex 21	1	2	0	0	0	1
22 : Ex 22	2	0	1	0	0	0
23 : Ex 23	0	3	2	0	5	5
24 : Ex 24	0	0	1	1	1	1
25 : Ex 25	1	1	0	0	0	0
26 : Ex 26	0	4	1	1	5	8

Table A2.8: Intelligent Labels for Examples

Example Intelligent Labels
Ex01_GF_ARCON_EMS_0_0_0_EN_TDM_0_0_0
Ex02_GM_ARCON_EMS_TS_0_L_EHPNNS_TDM_0_0_0
Ex03_GM_ARSH_EMA_0_0_0_EHPNNLP_0_0_0_0
Ex04_GM_ARCON_EMN_0_R_0_EHPSNNN_TD_0_0_0
Ex05_GF_ARSH_EMA_0_0_0_EHPNNLP_TD_C_0_0
Ex06_GM_ARW_EMA_0_R_0_EHPNNN_TD_0_0_BC
Ex07GF_ARCON_EMH_0_0_0_EHPNNS_0_0_F_0
Ex08_GM_ARSH_EMA_TS_0_L_EHPNN_TD_0_F_0
ExS09_GM_ARCON_EMH_0_0_0_EHPNN_TD_0_0_0
Ex10_GM_ARCON_EMH_0_0_0_EHPNN_0_0_F_0
Ex11_GF_ARCON_EMH_0_R_0_EHPLP_TD_0_0_0
Ex12_GF_ARW_EMN_0_0_L_EHPS_TDM_0_0_0
Ex13_GF_ARCON_EMS_0_R_0_0_0_0_0_0

Ex14_GM_ARCON_EMS_0_0_0_EN_TDM_0_0_0
Ex15_GM_ARW_EMN_TS_0_0_EN_0_0_0_0
Ex16_GF_ARCON_EMH_0_0_0_EHPNN_TD_0_0_0
Ex17_GM_ARCON_EMH_0_0_L_EHPNN_TD_0_0_0
Ex18_GM_ARSH_EMA_0_0_0_EHPN_0_0_0_0
Ex19_GM_ARCON_EMS_0_0_0_EHPNN_0_0_0_BC
Ex20_GF_ARSH_EMH_0_0_0_EHPN_TD_0_F_0
ExS21_GM_ARCOON_EMS_0_0_L_EHPNNNLP_TD_0_0_0
Ex22_GM_ARCON_EMH_0_0_0_EHPNNN_0_0_0_0
Ex23_GM_ARCON_EMN_TS_R_0_EHPNNLP_TDM_0_0_0
ExS24_GM_ARCON_EMN_0_0_0_ELP_0_0_F_0
Ex25_GF_ARSH_EMA_0_0_L_ENNN_TD_0_0_0
Ex26_GF_ARSH_EMH_TS_0_0_0_TD_0_0_0

Table A2.9: Intelligent Labels Key

Intelligent Labels Key
Ex + number = Example
GF and GM = Female or Male voice
AR = Delivery (CON = conversational; SH = shouting; W = whisper)
EM = Emotion (S = sad; A = Angry; N = neutral; H = happy
TS = Time-stretch (TS = yes; 0 = no)
R = Reversed (R = yes; 0 = no)
L = Loop (L = yes; 0 = no)
E = Equalisation (N = node; HP = high-pass; S = shelf; LP = low-pass)
TD = Tape delay (TD = tape delay; TDM = tape delay with modulation; 0 = no)
C = Compression (C = yes; 0 = no)
F = Flanger (F = yes; 0 = no)
BC = Bit crush (BC = yes; 0 = no)

Table A2.10: Correlation Chart between all Examples with a Female Source.

Correlations

			01:GF:ARCON:EMS:0:0:0:EN:TDM:0:0:0	05:GF:ARSH:EMA:0:0:0:EHPNNLP:TD:C:0:0	07:GF:ARCON:EMH:0:0:0:EHPNNS:0:0:F:0	11:GF:ARCON:EMH:0:R:0:EHPLP:TD:0:0:0	12:GF:ARW:EMN:0:0:L:EHPS:TDM:0:0:0	13:GF:ARCON:EMS:0:R:0:0:0:0:0	16:GF:ARCON:EMH:0:0:0:EHPNN:TD:0:0:0	20:GF:ARSH:EMH:0:0:0:EHPN:TD:0:F:0	25:GF:ARSH:EMA:0:0:L:ENNN:TD:0:0:0	26:GF:ARSH:EMH:TS:0:0:0:TD:0:0:0
Spearman's rho	01:GF:ARCON:EMS:0:0:0:EN:TDM:0:0:0	Correlation Coefficient	1,000	0,082	-0,108	0,388	0,362	0,487	-0,168	0,433	0,189	.579*
		Sig. (2-tailed)		0,781	0,712	0,170	0,203	0,077	0,565	0,122	0,518	0,030
		N	14	14	14	14	14	14	14	14	14	14
	05:GF:ARSH:EMA:0:0:0:EHPNNLP:TD:C:0:0	Correlation Coefficient	0,082	1,000	0,421	0,302	0,377	-0,141	0,523	0,274	0,054	0,312
		Sig. (2-tailed)	0,781		0,134	0,294	0,184	0,630	0,055	0,343	0,854	0,277
		N	14	14	14	14	14	14	14	14	14	14
	07:GF:ARCON:EMH:0:0:0:EHPNNS:0:0:F:0	Correlation Coefficient	-0,108	0,421	1,000	0,101	-0,121	0,514	0,449	0,426	-0,313	-0,023
		Sig. (2-tailed)	0,712	0,134		0,731	0,681	0,060	0,108	0,129	0,276	0,938
		N	14	14	14	14	14	14	14	14	14	14
	11:GF:ARCON:EMH:0:R:0:EHPLP:TD:0:0:0	Correlation Coefficient	0,388	0,302	0,101	1,000	0,025	0,332	0,214	.647*	-0,130	.738**
		Sig. (2-tailed)	0,170	0,294	0,731		0,933	0,247	0,462	0,012	0,658	0,003
		N	14	14	14	14	14	14	14	14	14	14
	12:GF:ARW:EMN:0:0:L:EHPS:TDM:0:0:0	Correlation Coefficient	0,362	0,377	-0,121	0,025	1,000	-0,133	0,162	0,133	.680*	0,471
		Sig. (2-tailed)	0,203	0,184	0,681	0,933		0,651	0,580	0,651	0,007	0,089
		N	14	14	14	14	14	14	14	14	14	14
	13:GF:ARCON:EMS:0:R:0:0:0:0:0	Correlation Coefficient	0,487	-0,141	0,514	0,332	-0,133	1,000	0,082	0,371	-0,227	0,349
		Sig. (2-tailed)	0,077	0,630	0,060	0,247	0,651		0,781	0,192	0,435	0,221
		N	14	14	14	14	14	14	14	14	14	14
	16:GF:ARCON:EMH:0:0:0:EHPNN:TD:0:0:0	Correlation Coefficient	-0,168	0,523	0,449	0,214	0,162	0,082	1,000	0,370	0,279	0,087
		Sig. (2-tailed)	0,565	0,055	0,108	0,462	0,580	0,781		0,193	0,334	0,766
		N	14	14	14	14	14	14	14	14	14	14
	20:GF:ARSH:EMH:0:0:0:EHPN:TD:0:F:0	Correlation Coefficient	0,433	0,274	0,426	.647*	0,133	0,371	0,370	1,000	0,140	0,403
		Sig. (2-tailed)	0,122	0,343	0,129	0,012	0,651	0,192	0,193		0,633	0,154
		N	14	14	14	14	14	14	14	14	14	14
	25:GF:ARSH:EMA:0:0:L:ENNN:TD:0:0:0	Correlation Coefficient	0,189	0,054	-0,313	-0,130	.680*	-0,227	0,279	0,140	1,000	0,144
		Sig. (2-tailed)	0,518	0,854	0,276	0,658	0,007	0,435	0,334	0,633		0,623
		N	14	14	14	14	14	14	14	14	14	14
	26:GF:ARSH:EMH:TS:0:0:0:TD:0:0:0	Correlation Coefficient	.579*	0,312	-0,023	.738**	0,471	0,349	0,087	0,403	0,144	1,000
		Sig. (2-tailed)	0,030	0,277	0,938	0,003	0,089	0,221	0,766	0,154	0,623	
		N	14	14	14	14	14	14	14	14	14	14

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

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

Table A2.11: Correlation Chart between all Examples with Male Source.

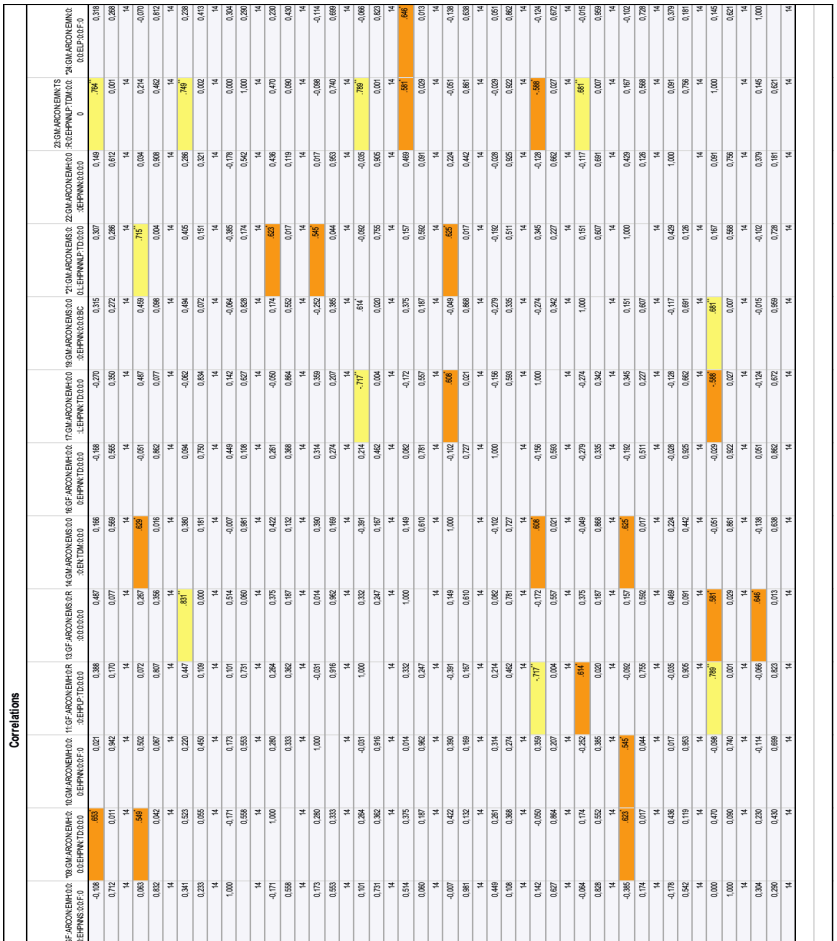
Correlations

			02:GMARCONEMST50L: EHPNNSTDM0/0/0	03:GMARSHEMA0/0/0/EHP NNLP0/0/0/0	04:GMARCONEMV1/0/E: HPNNNTD0/0/0	06:GMARVHEMA0/R/0/EHP NNNTD0/0/BC	08:GMARSHEMATS0/LEH PNNTD0/0/0	09:GMARCONEMH1/0/0: EHPNNTD0/0/0	10:GMARCONEMH1/0/E: HPNN0/0/F/0	14:GMARCONEMS0/0/E: NTDM0/0/0	15:GMARVHEMATS0/EN: 0/0/0/0	17:GMARCONEMH1/0/E: HPNNTD0/0/0	18:GMARSHEMA0/0/EHP NNNTD0/0/BC	19:GMARCONEMS0/0/E: HPNN0/0/0/BC	21:GMARCONEMS0/0/E: HPNNNLPTD0/0/0	22:GMARCONEMH1/0/E: HPNNN0/0/0/0	23:GMARCONEMNTSR/E: EHPNNLPTDM0/0	24:GMARCONEMN1/0/0: ELPN0/F/0
Spearman's rho	02:GMARCONEMST50L: EHPNNSTDM0/0/0	Correlation Coefficient	1.000	-0.198	0.305	0.090	0.458	.549*	.502*	.829**	0.236	0.487	0.264	.715*	0.034	0.214	-0.070
		Sig. (2-tailed)		0.498	0.285	0.758	0.099	0.042	0.067	0.016	0.416	0.077	0.201	0.098	0.004	0.462	0.812
		N	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14
	03:GMARSHEMA0/0/0/EHPNNLP0/0/0	Correlation Coefficient	-0.198	1.000	-0.158	-0.510	-0.376	-0.349	0.005	0.124	-0.510	0.457	0.063	-0.384	-0.266	-0.302	-0.379	0.129
		Sig. (2-tailed)	0.498		0.589	0.062	0.185	0.222	0.987	0.672	0.063	0.100	0.830	0.175	0.359	0.293	0.181	0.660
		N	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14
	04:GMARCONEMN1/0/R/0/EHPNNNTD0/0/0	Correlation Coefficient	0.305	-0.158	1.000	0.299	.823**	0.323	0.220	0.380	0.378	-0.062	0.163	0.494	0.405	0.286	.749**	0.238
		Sig. (2-tailed)	0.285	0.589		0.298	0.017	0.655	0.450	0.181	0.183	0.834	0.577	0.072	0.151	0.321	0.002	0.413
		N	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14
	06:GMARVHEMA0/R/0/EHPNNNTD0/0/BC	Correlation Coefficient	0.090	-0.510	0.299	1.000	0.167	-0.043	-0.244	-0.067	0.300	-0.271	0.084	0.473	0.124	0.224	0.475	0.106
		Sig. (2-tailed)	0.758	0.062	0.296		0.568	0.885	0.400	0.820	0.298	0.349	0.776	0.086	0.672	0.441	0.086	0.718
		N	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14
	08:GMARSHEMATS0/LEH PNNTD0/F/0	Correlation Coefficient	0.458	-0.376	.823**	0.167	1.000	.595*	0.529	0.455	0.217	0.010	-0.434	0.165	.894**	0.220	0.473	-0.302
		Sig. (2-tailed)	0.099	0.185	0.017	0.568		0.035	0.052	0.102	0.457	0.973	0.121	0.573	0.006	0.451	0.088	0.293
		N	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14
	09:GMARCONEMH1/0/0/EHPNNTD0/0/0	Correlation Coefficient	.549*	-0.349	0.323	-0.043	.595*	1.000	0.286	0.422	0.160	-0.050	-0.192	0.174	.823**	0.436	0.470	0.230
		Sig. (2-tailed)	0.042	0.222	0.055	0.885	0.035		0.333	0.132	0.584	0.864	0.510	0.552	0.017	0.119	0.090	0.430
		N	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14
	10:GMARCONEMH1/0/E/EHPNN1/0/F/0	Correlation Coefficient	0.502	0.005	0.220	-0.244	0.529	0.286	1.000	0.390	0.102	0.359	-0.169	-0.252	.545*	0.017	-0.098	-0.114
		Sig. (2-tailed)	0.067	0.987	0.450	0.400	0.052	0.333		0.169	0.728	0.207	0.564	0.385	0.044	0.953	0.740	0.699
		N	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14
	14:GMARCONEMS0/0/E/EN:TDM0/0/0	Correlation Coefficient	.829**	0.124	0.380	-0.067	0.455	0.422	0.390	1.000	-0.362	.808**	-0.186	-0.049	.825**	0.224	-0.051	-0.138
		Sig. (2-tailed)	0.016	0.672	0.181	0.820	0.102	0.132	0.169		0.204	0.021	0.525	0.868	0.017	0.442	0.861	0.638
		N	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14
	15:GMARVHEMATS0/EN:0/0/0/0	Correlation Coefficient	0.236	-0.510	0.378	0.300	0.217	0.160	0.102	-0.362	1.000	-0.439	0.425	0.521	-0.156	-0.316	.584*	0.053
		Sig. (2-tailed)	0.416	0.063	0.183	0.298	0.457	0.584	0.728	0.204		0.116	0.130	0.056	0.593	0.271	0.028	0.857
		N	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14
	17:GMARCONEMH1/0/L:EHPNN:TD0/0/0	Correlation Coefficient	0.487	0.457	-0.062	-0.271	0.010	-0.500	0.399	.806**	-0.439	1.000	0.140	-0.274	0.345	-0.128	.586*	-0.124
		Sig. (2-tailed)	0.077	0.100	0.834	0.349	0.973	0.864	0.207	0.021	0.116		0.633	0.342	0.227	0.662	0.027	0.672
		N	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14
	18:GMARSHEMA0/0/0/EHPNN1/0/0/0	Correlation Coefficient	0.264	0.063	0.163	0.084	-0.434	-0.192	-0.169	-0.186	0.425	0.140	1.000	.592*	-0.231	-0.392	0.123	0.211
		Sig. (2-tailed)	0.201	0.830	0.577	0.776	0.121	0.510	0.564	0.525	0.130	0.633		0.026	0.426	0.166	0.676	0.469
		N	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14
	19:GMARCONEMS0/0/0/EHPNN1/0/0/BC	Correlation Coefficient	0.459	-0.384	0.494	0.473	0.165	0.174	-0.292	-0.049	0.521	-0.274	.592*	1.000	0.151	-0.117	.881**	-0.015
		Sig. (2-tailed)	0.098	0.175	0.072	0.088	0.573	0.552	0.385	0.868	0.056	0.342	0.026		0.607	0.691	0.007	0.959
		N	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14
	*21:GMARCONEMS0/0/0/ELEHPNNNLPTD0/0/0	Correlation Coefficient	.715**	-0.266	0.405	0.124	.894**	.823**	.545*	.825**	-0.156	0.345	-0.231	0.151	1.000	0.429	0.167	-0.102
		Sig. (2-tailed)	0.004	0.359	0.151	0.672	0.006	0.017	0.044	0.017	0.593	0.227	0.426	0.607		0.126	0.568	0.729
		N	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14
	22:GMARCONEMH1/0/0/EHPNN1/0/0/0	Correlation Coefficient	0.034	-0.302	0.286	0.224	0.220	0.436	0.017	0.224	-0.316	-0.128	-0.392	-0.117	0.429	1.000	0.091	0.379
		Sig. (2-tailed)	0.908	0.293	0.321	0.441	0.451	0.119	0.953	0.442	0.271	0.662	0.166	0.691	0.126		0.756	0.181
		N	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14
	23:GMARCONEMNTSR/E:EHPNNLPTDM0/0/0	Correlation Coefficient	0.214	-0.379	.749**	0.475	0.473	0.470	-0.098	-0.051	.584*	-.586*	0.123	.881**	0.167	0.091	1.000	0.145
		Sig. (2-tailed)	0.462	0.181	0.002	0.086	0.088	0.090	0.740	0.861	0.028	0.027	0.676	0.007	0.568	0.756		0.621
		N	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14
	*24:GMARCONEMN1/0/0/ELPN0/F/0	Correlation Coefficient	-0.070	0.129	0.238	0.106	-0.302	0.230	-0.174	-0.138	0.053	-0.124	0.211	-0.015	-0.102	0.379	0.145	1.000
		Sig. (2-tailed)	0.812	0.660	0.413	0.718	0.293	0.430	0.699	0.638	0.857	0.672	0.469	0.959	0.728	0.181	0.621	
		N	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14

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

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

Table A2.12: Correlation chart between Examples with a Conversational Delivery



Correlations

		01:GF-ARCONEMB.0.0.1:ENTDM.0.0.0	02:GM-ARCONEMB.0.0.1:ENPNKT.D0.0.0	04:GM-ARCONEMB.TS.0.0.1:ENPNNKT.D0.0.0	05:GF-ARCONEMB.0.0.1:ENPNNKS.0.0.F.0	09:GM-ARCONEMB.0.0.1:ENPNKT.D0.0.0	10:GM-ARCONEMB.0.0.1:ENPNKS.0.0.F.0	11:GF-ARCONEMB.0.0.1:EPLP.T.D0.0.0	13:GF-ARCONEMB.0.0.1:EPLP.T.D0.0.0	14:GM-ARCONEMB.0.0.1:EN.TDM.0.0.0	16:GF-ARCONEMB.0.0.1:EPNKT.D0.0.0	17:GM-ARCONEMB.0.0.1:EPNKT.D0.0.0	19:GM-ARCONEMB.0.0.1:EPNKS.0.0.F.0	21:GM-ARCONEMB.0.0.1:EPNNKP.T.D0.0.0	22:GM-ARCONEMB.0.0.1:EPNNKS.0.0.F.0	23:GM-ARCONEMB.TS.0.0.1:EPNNKP.T.D0.0.0	24:GM-ARCONEMB.0.0.1:EP.1.0.F.0
Spearman's rho	01:GF-ARCONEMB.0.0.1:ENTDM.0.0.0	Correlation Coefficient	1,000	,261	,498*	,076	,712	,011	,942	,170	,077	,589	,595	,330	,272	,286	,612	,149	,784*	,538
		Sig. (2-tailed)	.	,367	,016	,795	,002	,969	,000	,560	,793	,027	,026	,248	,346	,320	,020	,611	,001	,047
		N	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14
	02:GM-ARCONEMB.TS.0.0.1:EPNNKS.TD.M.0.0.0	Correlation Coefficient	,261	1,000	,305	,083	,546*	,532	,072	,267	,826**	-,051	,487	,459	,715**	,034	,214	-,070		
		Sig. (2-tailed)	,367	.	,286	,832	,042	,067	,807	,356	,016	,862	,077	,098	,004	,906	,462	,812		
		N	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14		
	04:GM-ARCONEMB.0.0.1:EPNNKS.TD.0.0.0	Correlation Coefficient	,819**	,305	1,000	,341	,323	,220	,447	,831**	,380	,094	-,062	,494	,405	,286	,749**	,228		
		Sig. (2-tailed)	,016	,286	.	,233	,055	,450	,109	,000	,181	,750	,834	,072	,151	,321	,002	,413		
		N	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14		
	07:GF-ARCONEMB.0.0.0.1:EPNNKS.0.0.F.0	Correlation Coefficient	-,109	,063	,341	1,000	-,171	,173	,101	,514	-,007	,449	,142	-,064	-,365	-,178	,000	,304		
		Sig. (2-tailed)	,712	,832	,233	.	,558	,553	,731	,060	,981	,106	,627	,828	,174	,542	1,000	,290		
		N	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14		
	09:GM-ARCONEMB.0.0.0.1:EPNNKT.D.0.0.0	Correlation Coefficient	,853**	,546*	,523	-,171	1,000	,260	,294	,375	,422	,261	-,050	,174	,825**	,436	,470	,220		
		Sig. (2-tailed)	,011	,042	,055	,558	.	,333	,260	,187	,132	,368	,864	,552	,077	,119	,086	,450		
		N	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14		
	10:GM-ARCONEMB.0.0.0.1:EPNNKS.0.0.F.0	Correlation Coefficient	,021	,502	,220	,173	,260	1,000	-,031	,014	,360	,314	,359	-,252	,545*	,017	-,088	-,114		
		Sig. (2-tailed)	,942	,067	,450	,553	,333	.	,916	,962	,207	,276	,209	,385	,044	,953	,740	,699		
		N	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14		
	11:GF-ARCONEMB.0.0.0.1:EPLP.T.D0.0.0	Correlation Coefficient	,388	,072	,447	,101	,294	-,031	1,000	,332	-,361	,214	-,717**	,614	-,092	-,035	,789**	-,066		
		Sig. (2-tailed)	,170	,807	,109	,731	,362	,916	.	,247	,167	,462	,004	,000	,755	,905	,001	,823		
		N	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14		
	13:GF-ARCONEMB.0.0.0.1:EPLP.T.D0.0.0	Correlation Coefficient	,487	,287	,691*	,514	,375	,014	,332	1,000	,149	,092	-,172	,375	,157	,469	,981**	,645*		
		Sig. (2-tailed)	,077	,396	,000	,060	,187	,962	,247	.	,610	,761	,557	,187	,592	,091	,000	,013		
		N	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14		
	14:GM-ARCONEMB.0.0.0.1:EN.TDM.0.0.0	Correlation Coefficient	,589	,826**	,595	,360	,422	,360	-,391	,149	1,000	,610	-,102	,858**	,825**	,224	-,051	-,138		
		Sig. (2-tailed)	,026	,016	,026	,207	,132	,169	,167	,610	.	,727	,021	,868	,077	,442	,861	,638		
		N	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14		
	16:GF-ARCONEMB.0.0.0.1:EPNKT.D0.0.0	Correlation Coefficient	-,166	-,051	,094	,449	,261	,314	,214	,092	-,102	1,000	-,156	-,279	-,279	-,192	-,039	-,039		
		Sig. (2-tailed)	,566	,862	,750	,108	,368	,274	,462	,781	,727	.	,593	,335	,336	,511	,895	,922		
		N	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14		
	17:GM-ARCONEMB.0.0.0.1:EPNKT.D0.0.0	Correlation Coefficient	-,270	,487	-,062	,142	-,050	,359	-,717**	-,172	,826**	-,156	1,000	-,274	,345	-,128	-,888**	-,134		
		Sig. (2-tailed)	,350	,077	,834	,627	,864	,207	,004	,557	,021	,593	.	,342	,227	,662	,027	,672		
		N	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14		
	19:GM-ARCONEMB.0.0.0.1:EPNKS.0.0.F.0	Correlation Coefficient	,315	,459	,484	-,084	,174	-,252	,614*	,375	-,049	-,279	-,274	1,000	,151	-,117	,661**	-,015		
		Sig. (2-tailed)	,272	,098	,072	,828	,552	,385	,020	,187	,868	,335	,342	.	,607	,691	,007	,959		
		N	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14		
	21:GM-ARCONEMB.0.0.0.1:EPNNNP.T.D0.0.0	Correlation Coefficient	,307	,715**	,455	-,365	,825**	,545*	-,092	,157	,825**	-,192	,345	,151	1,000	,429	,167	-,102		
		Sig. (2-tailed)	,286	,004	,101	,174	,077	,044	,755	,592	,017	,511	,227	,607	.	,126	,568	,728		
		N	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14		
	22:GM-ARCONEMB.0.0.0.1:EPNNKS.0.0.F.0	Correlation Coefficient	,149	,034	,086	-,178	,436	,017	-,088	-,114	-,066	,469	,091	-,128	-,117	1,000	,091	,979**		
		Sig. (2-tailed)	,612	,906	,767	,542	,119	,953	,905	,699	,823	,091	,662	,691	,691	.	,756	,000		
		N	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14		
	23:GM-ARCONEMB.TS.0.0.1:EPNNKP.T.D0.0.0	Correlation Coefficient	,784**	,214	,749**	,000	,470	,044	-,088	,789**	,661**	-,092	-,368	,661**	,167	,091	1,000	,145		
		Sig. (2-tailed)	,001	,462	,002	1,000	,090	,880	,755	,001	,029	,861	,027	,007	,568	,756	.	,621		
		N	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14		
	24:GM-ARCONEMB.0.0.1:EP.1.0.F.0	Correlation Coefficient	,538	-,070	,228	,304	,220	-,114	-,066	,645*	-,138	,051	-,134	-,015	-,102	,979**	,145	1,000		
		Sig. (2-tailed)	,047	,812	,413	,290	,450	,699	,823	,013	,638	,862	,672	,959	,728	,000	,621	.		
		N	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14		

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

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

Table A2.13: Correlation Chart between Examples with Shouting Delivery.

Correlations

			03:GM:ARSH:EMA:0:0:0:EHPNNLP:0:0:0:0	05:GF:ARSH:EMA:0:0:0:EHPNNLP:TD:C:0:0	08:GM:ARSH:EMA:TS:0:L:EHPNN:TD:0:F:0	18:GM:ARSH:EMA:0:0:0:EHPN:0:0:0:0	20:GF:ARSH:EMH:0:0:0:EHPN:TD:0:F:0	25:GF:ARSH:EMA:0:0:L:ENNN:TD:0:0:0	26:GF:ARSH:EMH:TS:0:0:0:TD:0:0:0
Spearman's rho	03:GM:ARSH:EMA:0:0:0:EHPNNLP:0:0:0:0	Correlation Coefficient	1,000	0,015	-0,376	0,063	-0,103	0,200	-,598
		Sig. (2-tailed)		0,960	0,185	0,830	0,725	0,492	0,024
		N	14	14	14	14	14	14	14
	05:GF:ARSH:EMA:0:0:0:EHPNNLP:TD:C:0:0	Correlation Coefficient	0,015	1,000	-0,089	0,350	0,274	0,054	0,312
		Sig. (2-tailed)	0,960		0,761	0,220	0,343	0,854	0,277
		N	14	14	14	14	14	14	14
	08:GM:ARSH:EMA:TS:0:L:EHPNN:TD:0:F:0	Correlation Coefficient	-0,376	-0,089	1,000	-0,434	0,351	,581*	0,497
		Sig. (2-tailed)	0,185	0,761		0,121	0,218	0,029	0,070
		N	14	14	14	14	14	14	14
	18:GM:ARSH:EMA:0:0:0:EHPN:0:0:0:0	Correlation Coefficient	0,063	0,350	-0,434	1,000	-0,040	-0,521	0,246
		Sig. (2-tailed)	0,830	0,220	0,121		0,892	0,056	0,396
		N	14	14	14	14	14	14	14
	20:GF:ARSH:EMH:0:0:0:EHPN:TD:0:F:0	Correlation Coefficient	-0,103	0,274	0,351	-0,040	1,000	0,140	0,403
		Sig. (2-tailed)	0,725	0,343	0,218	0,892		0,633	0,154
		N	14	14	14	14	14	14	14
	25:GF:ARSH:EMA:0:0:L:ENNN:TD:0:0:0	Correlation Coefficient	0,200	0,054	,581*	-0,521	0,140	1,000	0,144
		Sig. (2-tailed)	0,492	0,854	0,029	0,056	0,633		0,623
		N	14	14	14	14	14	14	14
	26:GF:ARSH:EMH:TS:0:0:0:TD:0:0:0	Correlation Coefficient	-,598*	0,312	0,497	0,246	0,403	0,144	1,000
		Sig. (2-tailed)	0,024	0,277	0,070	0,396	0,154	0,623	
		N	14	14	14	14	14	14	14

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

Table A2.14: Correlation Chart between Examples with Whispered Delivery.

Correlations

			06:GM:ARW:EMA:0:R:0:EHPNNN:TD:0:0:BC	12:GF:ARW:EMN:0:0:L:EHPS:TDM:0:0:0	15:GM:ARW:EMN:TS:0:0:EN:0:0:0:0
Spearman's rho	06:GM:ARW:EMA:0:R:0:EHPNNN:TD:0:0:BC	Correlation Coefficient	1,000	-0,143	0,300
		Sig. (2-tailed)		0,625	0,298
		N	14	14	14
	12:GF:ARW:EMN:0:0:L:EHPS:TDM:0:0:0	Correlation Coefficient	-0,143	1,000	0,249
		Sig. (2-tailed)	0,625		0,391
		N	14	14	14
	15:GM:ARW:EMN:TS:0:0:EN:0:0:0:0	Correlation Coefficient	0,300	0,249	1,000
		Sig. (2-tailed)	0,298	0,391	
		N	14	14	14