UNIVERSITY OF NEW ENGLAND

Voices Within Voices: Developing a New Analytical Approach to Vocal Timbre by Examining the Interplay of Emotionally Valenced Vocal Timbres and Emotionally Valenced Lyrics

A Dissertation submitted by

Kristal L. Spreadborough, BMus(Hons), AMusA

For the award of Doctor of Philosophy

23rd March, 2018

Abstract

Aims/Goals

This thesis presents a new analytical technique for vocal timbre based on the hypothesis that emotion expressed in vocal timbre impacts emotional perception of lyrics.

Background information

Vocal timbre is a highly salient musical feature that, arguably, contributes significantly to our emotional experience of a song. Despite this, analytical techniques for vocal timbre remain in their infancy. Today, this is changing as technological developments increasingly allow for vocal timbre to be preserved and studied in a systematic way. The present research capitalises on these developments, using them to facilitate the examination of how emotional vocal timbres impact emotional perception of lyrics.

Methodology

Since there exists little empirical research on the hypothesis which underlies this analytical technique, and since the experience of vocal timbre could be considered highly subjective, it was necessary to first experimentally test if/how vocal timbre impacts lyric perception. To this end, a reception test was conducted to examine whether vocal timbre on its own has emotional valence, and whether this emotional valence is salient enough to impact emotional perception of words. Results from this test supported the hypothesis, showing that participants were significantly less accurate at identifying the emotional valence of words when these words were sung with a mismatched emotional vocal timbre.

The analytical technique itself is multilayered. First, the recording is taken as the basis of analysis. Then, the vocal timbre is described, and its emotional valence is assessed, through Vocal Timbre Features (a system, inspired by the work of van Leeuwen (1999) defined and developed to aid in describing vocal timbre and, potentially, categorising its emotional valence). Observations made by aurally detecting and annotating the Vocal Timbre Features can be confirmed visually through spectrographs. The synergies between emotions identified in the vocal timbre and that conveyed through lyrics can then be assessed using adapted diagrammatic vocabulary sets (inspired by the work of Dennis Smalley (1986, 1997)).

Conclusions

In summary, this thesis presents a new analytical technique that allows one to analyse vocal timbre in terms of its emotional meaning, and in terms of how this emotional meaning impacts emotional perception of lyrics. It also offers a framework through which one may conduct efficient, aurally based, analyses of vocal timbre more generally. This thesis has also shown that the experience of emotion in vocal timbre, and its impact on lyric perception, may be similar across listeners (i.e., intersubjective).

Certification of Dissertation

I certify that the ideas, experimental work, results, analyses, software and conclusions reported in this dissertation are entirely my own effort, except where otherwise acknowledged. I also certify that the work is original and has not been previously submitted for any other award, except where otherwise acknowledged.

Signature of Candidate   Date

ENDORSEMENT

Signature of Supervisor/s   Date

Signature of Supervisor/s   Date

Acknowledgements

As I sit to write these acknowledgments, and reflect on the journey that has led me to PhD completion, I realise how very fortunate I am to be part of such an encouraging and inclusive environment–both within the University of New England and beyond. I am humbled by the support that has been given to me over the last three years from the academic community, administrative staff, and my personal networks.

Foremost among these supporters are my supervisors, to whom I will be forever indebted: Dr Inés Antón-Méndez and Dr. Donna Hewitt. The immeasurable amount of time, energy, and guidance they have invested has given this project wings. Although these few words cannot adequately express my gratitude, special thanks must be given to my principal supervisor Inés Antón-Méndez for her forbearance and sage advice at every turn. Many thanks also go to my past supervisor Dr. Jenny Game-Lopata for her encouragement in the early days of this project.

I wish to acknowledge the contribution of the Australian Government for providing funding through the Australian Postgraduate Award. Special thanks are also given to the colleagues who read drafts of this manuscript and offered their constructive criticism. To the many administrative staff who work "behind the scenes" (and who let me into my office on more than one occasion when I had locked myself out) – a heartfelt thank you. I would also like to recognise the work of Phillipa Trelford for her excellent editorial assistance.

Support has come in many shapes from friends and family both near and far. To my parents, family, and friends, your kind words were more of a comfort

than you might know. To Melita, Kylie, and Alana – never underestimate the value of those coffees on the deck, cups of tea in the kitchen, and glasses of wine out the back with "the girls". They gave me much needed moments of respite. Finally, a heartfelt thanks goes to my partner Steve for being patient and kind, and for always being there.

Publications Arising From This Thesis

Spreadborough, K. L., & Anton-Mendez, I. (2018). It's not what you sing, it's how you sing it: How the emotional valence of vocal timbre influences listeners' emotional perception of words. *Psychology of Music*. https://doi:10.1177/0305735617753996

Table of Contents

List of Figures

List of Tables

1 General Introduction and Overview

1.1 The Emotional Impact of Vocal Timbre in Song

Sound plays an important role in our everyday lives and communications (E. C. Blake & Cross, 2015, pp. 81–82). Listeners can derive a variety of meanings from a single sound. For example, a haphazard pedestrian crossing a busy street without looking could only jump out of the way of oncoming traffic by knowing the direction and speed of the approaching cars; a zoo keeper may be able to assume a lion’s intent to attack by listening to its growl; a judge may gauge a criminal’s level of remorse by the tone of their voice. In these situations, sounds play a significant role in the expression and perception of meaning.

Musical sounds too may convey meaning. For example, research has shown that listeners derive meaning about phrase structure and pulse from harmony and rhythm (see Chapter 3, section 3.4.1), and that timbres alone can carry meaning about instrument identity (See Chapter 3, section 3.4.2). The impact of sound, then, is not limited only to one’s *everyday* experiences, but also extends to the interpretation of more *abstract* experiences.

The spoken voice in particular is a sound source that can be highly evocative, serving as a vehicle for linguistic expression. Sayings such as “it’s not what you said, it’s how you said it” and “don’t take that tone with me,” are examples of how the way something is said can be just as important as what is being said (see Chapter 3, section 3.3). Many studies have found that the spoken voice provides important cues about a person's emotional state (Aucouturiera et al., 2015; Heidemann, 2016; Krestar & McLennan, 2013; Lewis, 2000; Nygaard & Lundersl, 2002; Nygaard & Quees, 2008). The sound of the spoken voice is shown from both anecdotal and scientific evidence to be a particularly salient and expressive source of meaning.[1]

It is clear from the literature that the spoken voice is a good tool for conveying a speaker's emotion and transmitting meaning. The same may also be true of the *sung* voice. One reason to suppose the sung voice may be just as good at, and important for, conveying emotion and meaning may be because the expression of emotion through the spoken voice is the result of emotion's impact on the physical body—for example, the shape of the vocal tract, the use of air from the lungs. These physical characteristics may be an inherent part of communication and may be introduced unconsciously by the speaker. If it is the case that these bodily effects are so intertwined in the production of the spoken voice, then similar features which are a natural and unintended consequence of a speaker's emotional state may also manifest themselves in singing—after all, both modalities are produced by the same instrument. Another reason may be that generally in both spoken and sung voice, the

[1] The term *meaning* is a multifaceted and complex one. This thesis takes a pragmatic approach to the definition of meaning. Meaning is taken as the result of the association between a symbol and a concept. So, for example, a bomb-raid siren (the symbol) means danger is approaching (the concept).

meaning of the words needs to be easily understood.[2] Since the spoken voice is such a well-developed (and, for most people, a well-practiced) system for conveying linguistic content, it seems reasonable to expect that the sung voice will take advantage of the mechanisms developed for speaking. Singing, then, is a sound source which merges the everyday experience of sound (through its connection to the spoken voice) with the abstract experience of sound (as it occurs in song).

Scholars have proposed that this may be precisely the reason for the potential of the sung voice to be so emotionally impactful: because music exaggerates the evocative qualities already inherent within the human voice. Serge Lacasse, for example, argues that in popular song, the sung voice not only draws expressive and communicative qualities from spoken voice, but also exaggerates them:

> ...popular music codes and meanings are, arguably, deeply rooted in the everyday. More specifically, popular singing acquires most of its inspiration from everyday speech. According to the British singer-songwriter Leon Rosselson, "[t]he language of song, like the language of drama, is not a literary language; it embraces the idioms and rhythms

[2] There are, of course, some forms of music in which the words of the sung voice, while being essential to the song, are not easily understood. In some cases, the sound of the voice itself takes priority over the sung words. However, this is not the case in the songs that are to be the focus of this thesis. This is discussed in more detail in section 1.3.

> of everyday speech while looking for ways of enriching that language.” (Lacasse, 2010, pp. 226–227)

That reliance on and exaggeration of spoken vocal features in music more generally may enhance music’s emotional expressiveness has also been shown through scientific lines of enquiry. For example, in the work of Patrick N. Juslin and others (Juslin & Laukka, 2004; Juslin & Timmers, 2011).

Since the development of recordings, the expressiveness of the sung voice has, for the first time, been able to be preserved and played back repeatedly if desired. There are a number of musical genres which not only take advantage of recordings, but have, historically, relied on them as a fundamental tool for creation and dissemination. Popular music (of which the popular songs that are to be the focus of this thesis are a subset) is one such genre. Because of this music’s reliance on such recording-centred methods of creation and preservation, it has been proposed that:

> In contemporary popular music sound matters more. Key singers and instrumentalists develop their own, immediately recognisable styles of singing and playing, and, thanks to recording, these can now become part of the language of music and be imitated and transformed by countless others. Saxophones can be soft and mellow, or tense and strident, like a hoarse whisper or a foghorn in the mist. Voices can be soft, smooth and well oiled, or rough, raspy and cracked. And singers as well as instrumentalists use a large repertoire of howls, wails, groans and other vocalisations. (van Leeuwen, 1999, p. 127)

In this way, vocal nuances such as breath and noise are not only by-products of vocal production, but rather they may now form part of a system of indicators used by performers to express meaning and emotion in popular music.

This thesis investigates how these vocal nuances (which contribute to vocal timbre) may convey emotion, and how they may impact emotional perception of sung words. If it is the case that vocal timbre is such a highly salient sound source that may impact lyric perception, one may ask why more has not been done on this topic. One key reason is that analytical techniques have, in general, not kept pace with changes in popular music (see, e.g., Middleton, 1993; Seeger, 1977; Tagg, 2000, for a general discussion). Further, in the past, vocal timbre has been primarily ephemeral in nature, existing only in the live performance. This has hindered vocal timbre analysis as musicological study has historically focused on musical features which can be transmitted through conventional scores and traditional notation techniques.

Today this situation is changing. Developments in recording and playback now allow vocal timbre, a musical feature which would have otherwise been lost, to be preserved and widely shared. Popular song (a subset of popular music more generally) is a genre in which recordings, playback devices, as well as a raft of other technological developments (such as amplification and production effects) play a central role in its creation and dissemination.

These same developments are also allowing for vocal timbre to be studied in new ways. For example, vocal timbre can now be studied directly from the recording, heard repeatedly through playback devices, visualised through spectrographs, and systematically investigated using interdisciplinary approaches. Additionally, vocal timbre can also now be manipulated to a greater extent through the use of recording and playback devices, amplification, and other processing effects, potentially increasing its impact on the listener. For example, the development of the microphone meant that singers could use a range of dynamic and stylistic variations without being overpowered by other instruments—they could even *whisper* to their audience and still be heard. In this way, technological developments have allowed for vocal timbre in popular songs to be engaged with and manipulated in new ways, foregrounding vocal utterances, and making the study and analysis of vocal timbre more feasible.

Clearly, vocal timbre is a rich and diverse feature of music which should be the subject of more targeted research and analysis. Its ephemeral nature, however, has until relatively recently made it more challenging to study and has contributed to the dearth of analytical techniques on vocal timbre. The goal of this thesis is to go some way to filling that void—to take an interdisciplinary approach to the study and exploration of vocal timbre, how emotion is perceived in vocal timbre, and vocal timbre's function within music.

1.2 Thesis Aims

The main aim of this thesis is to present a new analytical technique for vocal timbre analysis based on the hypothesis that emotion conveyed through vocal timbre impacts emotional perception of lyrics. Linked to the development of this new technique, this thesis also aims to:

- Review the multi-disciplinary evidence that vocal timbre is a salient musical feature and that it can be used by performers to express emotional meaning. This thesis examines literature from a broad range of relevant fields including musicology, psycholinguistics, psychology, phonomusicology, electroacoustic music study, and semiotics. Through surveying this literature it will be shown that humans draw on, and are impacted by, information from a wide range of nonlinguistic and para-linguistic[3] sources for communication, that vocal timbre does play an important role in communicating emotion, and that there is room for new analytical techniques for vocal timbre.
- Use reception tests to investigate whether the experience of emotion in vocal timbre, and the impact of vocal timbre on lyric perception, is intersubjective (i.e., similar across a cohort of listeners). Such tests are needed as there is little empirical research on if/how listeners perceive emotion in vocal timbre specifically and if/how this impacts emotional perception of lyrics. Using such testing to

[3] Para-linguistic refers to the information we derive from a speaker's output that is not linguistic. For example, their tone of voice or their facial expressions.

support the underlying principles of the analytical technique developed here allows the analyses of vocal timbre conducted in this thesis to be considered—not just the idiosyncratic experience of the analyser, but rather to represent the intersubjective experience of vocal timbre more generally.

- Investigate, through the use of the reception tests mentioned in the point above, what may make vocal timbre a salient and evocative sound source.

1.3 Scope

The epistemological and methodological background.

This thesis crosses boundaries between the disciplines of musicology and psychology. It brings together popular music studies and psycholinguistics in order to create a new analytical technique for vocal timbre. It is essentially abductive in its approach. The already existing evidence points to the possibility that vocal qualities do impact perception, but how this translates to vocal timbre is yet to be explored. In addition to examining the existing literature, this thesis will also take a somewhat positivist approach, conducting reception tests to examine the relationship between emotional vocal timbres and emotional lyrics. However, a purely positivist approach is not completely sufficient for understanding this aspect of vocal timbre. For this reason, a critical approach is taken when defining vocal timbre in section 2.2. There, different modes of analysing vocal timbre will be discussed. Through the mixed approach of using quantitative data to support and supplement a musicological study, this thesis aims to achieve a reflexive methodology – it takes the existing literature on voice quality, vocal timbre, and popular music studies as a point of departure in exploring new ways to analyse and conceptualise vocal timbre.

Why emotion?

Emotion, rather than affect or mood, has been selected as the subject of investigation in this thesis as it is generally agreed that emotion relates to short lived "responses to situations that are perceived as relevant to an individual's current goals" (Gross, 2010, p. 212). In other words, "emotional episodes are elicited by something, are reactions to something, and are generally about something" (Ekkekakis, 2012, p. 232). This makes emotion distinct from other concepts such as mood, which can occur without any exposure to a stimulus, and affect, which is a kind of catch all phrase for emotion as well as mood and feeling (Gross, 2010; Ekkekakis, 2012). As will be theorised further below, vocal timbre may be related to a listener's "real life" experience of voice quality, and understanding a speaker's voice quality can be directly relevant to a listener's current goals. As such, vocal timbre should provide situational cues that have the power to elicit an emotional response from a listener. Since emotion is an immediate response to a situational cue, and since vocal timbre may be an especially salient situational cue because of its similarity to voice quality, emotion has been selected as the topic of study in this thesis.

Some theorists argue that there are evolutionary and biological reasons why some emotions are common among humans (Ekman, 1992; Johnson-laird & Oatley, 1992; Plutchik, 1991). This theory of “universal” emotions is important for this thesis as it seeks to not only develop a new analytical technique for vocal timbre, but to do so in a way that allows the results of an analysis to be considered more than just the idiosyncratic experience of the analyser. If emotions, and their triggers and *raisons d'être*, can go beyond an individual's idiosyncratic experience, then an analysis of emotions in vocal timbre may also go beyond the idiosyncratic experience of the analyser. According to Ekman, facial expressions act as signifiers of emotions (Ekman, 1999). In this thesis, I propose that the form of certain acoustic features within a vocal timbre may also act as signifiers of emotion (this is discussed in Part Two and Chapter Seven of this thesis) perhaps due in part to the way facial expressions alter the shape of the vocal tract and, as a result, affect vocal timbre.

Of all the possible emotions, the general emotions of happiness and sadness are used in this thesis because, broadly speaking, they appear in most theories of basic emotions (see Ortony & Turner, 1990, Table 1). However, in general, I have chosen not to model my investigation on any one particular *theory* of emotions for two reasons. First, the concept of what emotion *is* is constantly changing, and the idea that there exists a discrete set of basic human emotions is itself debated (see Introduction in Reevey, Ozer, and Ito, 2010, for an overview). Therefore, to base the present study on any set of basic human emotions would also require a judgment to be made about the nature of these theories of emotion, something which is neither a goal of this thesis, nor necessary for its the purposes.

Why popular vocal songs?

The analytical technique developed in this thesis is most relevant to those forms of popular music where the vocal line plays a key role in conveying meaning and emotion. For ease of reference, I will henceforth call these popular vocal songs (using popular vocal songs as a resource for studying timbre will be discussed in more detail in Chapter 4). Such popular vocal songs tend to be characterised as being genres and styles produced and performed from the early 20th century to the early 21st century predominantly by untrained musicians, and as being communicative within culturally and socially inclusive groups (Blacking, 1981, p. 6; Breen, 1987, p. 9; Lull, 1987). It is important for the analytical technique proposed in this thesis that the words of such songs be easily discernible. Therefore, this thesis will focus on popular vocal song categories where the vocal line is prominent or foregrounded, and where the lyrical content is also an important, and easily understandable, feature. In these categories, lyrics may be placed in the foreground of the listening experience as they may play a dominant role in conveying meaning, specifically emotional meaning, and contribute to the form of the song. Examining songs that foreground the vocal line in this way seems to be a logical starting place for examining the role of vocal timbre and the way it may contribute to the emotional meaning of a song.

1.4 Approach

To achieve the understanding necessary to develop an analytical technique which considers how emotional meaning is created by the interaction of both vocal timbre and words, this thesis will primarily draw on two broad approaches: music studies and psycholinguistics.

A musicological approach is used to address the popular vocal songs themselves. This approach underpins the methodology of the analytical technique developed in this thesis (and outlined in detail in Chapter 7). It is here that the musical text is dealt with.

A psycholinguistic approach is used to address the question of whether vocal timbre does in fact impact emotional perception of sung words and whether this occurs in an intersubjective way. That is, the analytical technique outlined in this thesis is based on the potential for vocal timbre to be salient enough to impact emotional perception of sung words. However, the emotional pull of the vocal timbre is not something that has been previously experimentally tested. As such, analysing vocal timbre in this way can be problematic as it is difficult to know whether the analyser's experience of vocal timbre is idiosyncratic, or shared with other listeners. Another consideration in this analytical technique is the impact of emotionally valenced[4] lyrics. As the analytical technique developed in this thesis aims to account for both vocal timbre and words, the emotive content of lyrics should also be considered.

[4] Emotional valence refers to "the extent to which an emotion is positive or negative" (Citron et al., 2014, para 3). It is concerned with an emotions inherent level of attractiveness or aversiveness.

To this end, reception tests are used to ascertain whether:

1. listeners perceive emotional meaning in vocal timbre;
2. emotion conveyed through vocal timbre impacts emotional perception of words; and
3. the experiences of emotion listed in points 1 and 2 occur in an intersubjective way.

By using reception tests to examine the model which underlies the analytical technique, this thesis builds on a growing trend to use scientific modes of enquiry to inform musicological study (see, for example, Lacasse, 2000; Lacasse & Lefrancois, 2008; Tagg, 1999, 2000, 2011, 2012; Tagg & Clarida, 2003).

1.5 Thesis Overview

This thesis is divided into three parts. Part I provides the theoretical background. One of the first things to establish in this background is the evidence that vocal timbre is in fact an important musical feature that is likely to be impacting our musical experience. This is addressed in Chapter 2. Given the need to establish empirically that listeners can perceive emotion in timbre and that this emotion impacts the emotional perception of lyrics in an intersubjective way, Chapter 3 reviews evidence that experimentally testing vocal timbre may yield meaningful results. If it is that vocal timbre seems like a promising musical element to analyse and test, one may ask why more has not been done in this area. This is addressed in Chapter 4. This chapter also presents the rationale/justification and groundwork for the development of a new analytical technique for vocal timbre.

In Part II, I present the results of the reception tests. These tests show that emotion conveyed through vocal timbre does seem to impact emotional perception of lyrics and does so in a similar way for a cohort of listeners. In other words, they suggest that the experience of emotion in vocal timbre can be considered intersubjective, and, therefore, that the analytical technique developed and implemented in this thesis is robust. This is addressed in Chapter 5, which describes the reception tests and their findings. Chapter 6 discusses a set of findings that could allow the systematic identification of emotional valence from a limited set of musical attributes.

In Part III of this thesis the new analytical technique of vocal timbre is introduced. The technique itself is described in Chapter 7. Chapter 8 presents an application of this technique to two popular vocal songs.

While the development of a new analytical technique for vocal timbre in popular vocal songs has been fruitful, much remains to be done to understand the role of vocal timbre in the context of an entire popular vocal song (i.e., when considering vocal timbre as it occurs with other instrumentation and musical elements) and to better understand how emotional meaning is conveyed through such songs. Chapter 9 outlines further avenues for research in these areas and offers a summary of the thesis.

Part I: Background

Sound never just “expresses” or “represents”, it always also, and at the same time, affects us. (van Leeuwen, 1991, p. 128)

We constantly inhabit the universe of the voices, we are continuously bombarded by voices, we have to make our daily way through a jungle of voices, and we have to use all kinds of machetes and compasses so as not to get lost. There are voices of other people, the voices of music, the voices of media, our own voices intermingled with the lot. All these voices are shouting, whispering, crying, caressing, threatening, imploring, seducing, commanding, pleading, praying, hypnotizing, confessing, terrorizing, declaring . . . (Dolar, 2006, p. 13)

So there sat popular music at the beginning of the 50s. No one got worked up about it; nobody talked about its social significance; nobody analysed it, discussed it, looked for double meanings. (Cole, 1970, pp. 10–11)

2 Vocal Timbre as an Important Musical Feature

2.1 Introduction

At the centre of this thesis is the idea that vocal timbre is an important musical parameter in popular vocal songs. One reason to suppose that this is the case is the way performers have taken advantage of technological developments to foreground timbral nuances. For example, the development of the “crooning” style of singing—a style characterised by a softer, more sentimental, approach to voice production—has only been made possible due to the developments in microphones in the 20th century. This style is exemplified in, for instance, Frank Sinatra’s “I’ve Got You Under My Skin” (Porter, 1956/1987, track 9) and Nat King Cole’s “Unforgettable” (Gordon, 1951/1954, track 1). However, although technology may allow vocal timbre to be used in new ways, timbre and vocal timbre have played a central role in music creation and performance well before the development of electronic amplification and audio recording.

This chapter will show that there is a long history of vocal timbre and timbre more broadly being considered important and meaningful features of music. Establishing this lineage lends credence to the idea that vocal timbre is an important musical parameter in popular vocal songs and, therefore, is worthy of analysis. That is, it supports the idea that timbre and vocal timbre have, throughout Western music history, always functioned as essential musical parameters, with popular vocal songs being one more type of Western music to take advantage of these features.

2.2 Defining Timbre and Vocal Timbre

Before discussing the role of timbre in music, it is important to define what is meant by the term. Timbre is the characteristic quality of a sound which allows us to distinguish it from other sounds. It is a multifaceted musical feature, and the difficulties in defining it reflect this.

There are various ways in which timbre is defined in acoustical studies. For example, the American National Standards Institute (ANSI) defines timbre as:

> …that attribute of auditory sensation in terms of which a listener can judge that two sounds, similarly presented and having the same loudness and pitch, are different. (Houtsma, 1997, p. 105)

This definition is problematic as it excludes sounds that are unpitched (Bregman, 1990, p. 92), yet which still have a discernible timbre—for example, sounds such as percussion, much of electroacoustic music and, particular to this thesis, vocal timbre manipulations such as screams, grunts, breaths, and so forth.

A more inclusive definition has been offered by Pratt and Doak in 1976: timbre "is that attribute of auditory sensation whereby a listener can judge that two sounds are dissimilar using any criteria other than pitch, loudness, or duration" (Pratt & Doak, 1976, as cited in Rossing, 1990, p. 125). However, rather than describing what timbre *is,* this definition describes what it is *not*.

Timbre can also be examined from a psychoacoustic perspective. From this perspective, timbre is strongly dependent on a sounds spectrum, as well as its envelope and frequency (Rossing, 1990, p. 80). The spectrum refers to how the sound's intensity is distributed as a function of frequency over time. Varying the strength of certain harmonics within the spectrum impacts the quality of a sound (Rossing, 1990, p. 127). The envelope refers to how a sound varies through time. For example, the speed at which a note begins (i.e., attack), what happens immediately after the attack (i.e., decay), how a note is prolonged (i.e., sustain) and the speed at which a note ends (i.e., release) all form part of the envelope. Frequency refers to the "number of vibrations per second" and is measured in hertz (Rossing, 1990, p. 31). Variations in the spectrum, envelope, and frequency of a sound produced by instruments and voices is what makes them so aurally distinctive.

A psychoacoustic approach to timbre is certainly useful in teasing out aspects of timbre as an acoustic phenomenon. In some lines of study, such as Fourier analysis, this approach is essential to one's understanding of timbre and to how one approaches its analysis. However, there is no consensus on a single psychoacoustic definition of timbre. Furthermore, in research such as that undertaken in this thesis, a psychoacoustic definition would not be entirely appropriate since, as Bregman observed (see above), current psychoacoustic definitions cannot always account for non-pitched aspects of timbre which are often part of the vocal line in popular vocal song. For these reasons, it may be useful to examine timbre on another level – a phenomenological one.

In this vein, Hajda, Kendall, Carterette, and Harshberger offer a broader and more flexible definition:

> Based on research findings and [previous] definitions . . . it is clear that timbre has two principal constituents: (1) It “conveys the identity of the instrument that produced it” (Butler, 1992, p. 238), and (2) It is representable by a palette or family of palettes (see Martens, 1985) in which tones from different sources can be related along perceptual dimensions. The first constituent is nominal or categorical in nature: the clarinet has a characteristic to its sound, regardless of the pitch, loudness, etc. The second constituent is a hybrid of categorical and ordinal organization: the clarinet is not nasal and is therefore differentiated from the oboe, which is nasal. (Hajda et al., 2004, p. 282)

Hajda et al.’s definition reflects not only a psychoacoustic understanding of timbre, but also the importance of how timbre is perceived by a listener. It allows for the timbre of the voice to be distinguished from timbres of other instruments. It also allows for non-pitched sounds to be included in the definition of timbre, and acknowledges that variation may be a key component of timbre. Such an approach is useful for this thesis as variation in the voice (e.g., breathy, throaty sounds, which may act as indicators of emotion) is commonplace in popular vocal songs. Other scholars have also acknowledged the importance of timbral variation as a key component of its expressiveness (see, for example, van Leeuwen, 1999, p. 127).

On this level, vocal timbre, particularly in popular vocal song, shares important similarities with paralinguistic aspects of speech. Paralinguistic features, as described by Roger Wescot, are “non-phonemic alterations of the pitch, stress, or tempo of ordinary speech, as in growling, shouting, or drawling” (1992, p. 30). Vocalisations of this kind have also been found to be important for conveying emotion and meaning in popular song. For example, Paul and Huron (2010) found that the “breaking voice” plays an essential role in conveying grief in Country songs. Similarities between musical timbre and paralinguistic cues have also been found in the expression of sarcasm, which appears to be reliably marked by nasality in both spoken and musical contexts (Plazak, 2010, as cited in Huron, 2015, p. 190). Parallels have also been found between the paralinguistic and timbral expression of sadness since, in both cases, sadness tends to be conveyed through darker timbres (Huron, 2015, p. 193). One reason that these paralinguistic features may be expressive is because they draw on phenomenological and semiotic relationships. For example, sadness is a low arousal emotional state. In this state, it is easiest to speak with a tone of voice which is darker in timbral colour. Therefore, a link between darker timbres and sadness may be forged. This link, it seems, is salient enough to carry over into musical expression. Because of these overlaps in paralinguistic and musical expression, the phenomenological and semiotic study of spoken language can inform our understanding of vocal timbre.

Based on this multi layered approach to vocal timbre, and using the structure set out in Hajda et al.'s (2004, p. 282) definition above, I now offer the following working definition of vocal timbre:

Vocal timbre is: (a) the characteristic sound of a singer's voice that differentiates it from other sound sources (and indeed from other singers), even when they share the same volume (dynamic), frequency (pitch), and time (duration), and; (b) a variety of pitched and unpitched vocal sounds (breathiness, stylised screams/cries, throaty sounds, etc.) derived from the way in which a single singer may use their body to produce sound for musical expression.

The definition of vocal timbre is in a state of constant flux, with definitions tending to be nuanced and specific to the context in which timbre is being studied. It is for this reason, I have put forth a working definition of timbre for the purpose of this study which takes into account the existing definitions in the literature.

2.3 Timbre and Vocal Timbre before the Western Common Practice Period

Music has been an important and meaningful musical parameter in many Western musical cultures. Ancient Greece and Rome, for example, have left many clues (artwork and other relics) about the nature of their musical landscapes, created by the kinds of instruments (and therefore their timbre) that would have been played. Remnants of early forms of notation even suggest what kinds of notes may have been placed together. These artefacts suggest that music (of which timbre is an important aspect) was an important element of ancient cultures and societies.

Indeed, it is known that "socially organised sounds such as music" did play a role in people's everyday lives, in rituals and significant events, and in learning and teaching (E. C. Blake & Cross, 2015, pp. 82–83). Therefore, because of the widespread use of music in everyday life, descriptions and references that indicate that vocal timbre and timbre are important, and potentially meaningful, musical features do exist. The Greek myth of the Sirens is one such example. This myth describes how the Sirens used their sweet voices to lure their victims to their deaths. Such mythology suggests that the voice had the potential to be an evocative musical parameter, and vocal timbre, as a salient feature of the sung voice, may have played a part in this.

References to timbre and vocal timbre can also be found throughout the first millennium. For example, in *Laws,* Plato asserts that emulating one instrument's timbre on another instrument can "corrupt" the listener: "Possessed by a frantic and unhallowed lust for pleasure they . . . actually imitated the sound of the flute on the harp" (as cited in Wess & Taruskin, 2008, p. 6). While this is not directly related to voice production and perception mechanisms, this observation does suggest that violating the connection between an instrument and its unique sound could result in the corruption of a listener and may suggest listeners perceived or recognised a meaningful relationship between an instrument and its timbre. Furthermore, according to Wess and Taruskin, Plato's discussion on hearing and speech is evidence of the connection between voice, music, and meaning—hearing and speech have been given to humans so that they may better understand music, which is important as music may be a means for understanding the universe at large (Wess & Taruskin, 2008, p. 8). Music may be linked more broadly to the voice as "so much of music", writes Plato, "is adapted to the sound of the voice" (as cited in Wess & Taruskin, 2008, p. 8).

Further examples of the connection between the voice, music, and meaning can be found in Quintilian's instructions for the training of oratoria. Quintilian calls for all oratoria to be trained in music so that they may more easily call on the inflections of the voice (as cited in Wess & Taruskin, 2008, pp. 10–11). This instruction suggests that a connection between music and the voice was observed, particularly in terms of drawing on music to facilitate emotionally meaningful vocalisations. Some texts on religion and religious ceremony also make reference to the act of singing being, at intermittent points in history, important for religious purposes. Saint Basil the Great (330 –379 A.D.) refers to the use of singing to subtly introduce lay people to biblical texts, the rationale being that a text sung was "softer" and more easily memorised (Wess & Taruskin, 2008, p. 21).

References to the quality of the sung voice can be found in the second millennium too. For example, in his 16th century text *Musica Choralis Deudsch*, Agricola discusses the characteristic qualities of different solmisation,[5] observing that certain syllables of solmisation seemed to have inherent vocal timbre qualities:

> Of the above-mentioned six voices, two are called b molles, namely ut and fa, for they are sung extremely mildly, gently, sweetly and softly. They are of one nature and character; therefore where the one may be sung, so may the other also be sung.
>
> Re and sol are called the middle or natural voices because they emit an average sound, not too mild or too clear [scharff].
>
> Mi and la are called durales, that is clear [scharff] and hard syllables. For they should and must be sung in a more manly and stronger [dapfferer] way than the b molles and naturals.
>
> This difference, when it is well noted and truly observed in singing, makes all melodies sweet and pleasing. (as cited in A. Smith, 2011, pp. 25–26)

Such observations suggest that the sound of the voice (vocal timbre) was seen as an important aspect of music performance. That the control of certain vocal sounds has the potential to make “all melodies sweet and pleasing” (as cited in A. Smith, 2011, pp. 25–26) shows that vocal timbre was not just a musical by-product, but was considered at the time to be musically meaningful within itself.

[5] Solmisation is, put simply, a system through which each note in a scale is given a unique syllable.

In general, the references to the manipulation of timbre and vocal timbre in the texts explored in this section suggest that these elements were considered important musical features. A recognition of their intrinsic musical value, and their salience as musical features, can be found throughout Western music history.

2.4 Timbre and Vocal Timbre in the Western Common Practice Period and Beyond

A similar interest in timbre and vocal timbre can be found in music from the Western common practice period. This can be seen throughout many instrumental and vocal works from this time. One obvious example can be found in the works of French composers from the 19th–20th centuries, which are notable for their impressionist-like musical writing, achieving new timbral qualities by placing timbres in new contexts within a piece (Pasler, 2001). Claude Debussy and Maurice Ravel are just two composers from this period who challenged conceptions of timbre's role in music, and composed for this musical feature in new ways (Pasler, 2001).

However, while timbre was increasingly being experimented with, it remained the case that it was not (could not in most cases be) documented or preserved, unlike other features of music such as melody or harmony. The only time that timbre could be encountered by listeners was in the live performance. The development of recording technology in the 20th century changed this, allowing timbre to be experienced “on demand” by anyone with access to a recording. Therefore, it is perhaps not surprising that such developments had a significant impact on the number of instances in which one sees timbre being discussed in a more *explicit and systematic* way.

The way in which performers learn music is one example of how recordings impacted our relationship with timbre. In addition to the notes and rhythms, the *sound of a singer or instrumentalist* can now be shared and studied:

> For performers of popular music, recordings have been especially valuable learning aids. The available scores do not always represent performances adequately, and they cannot easily indicate the timbres and sonic effects that musicians seek to develop. An aspiring rock guitarist once explained why he studied recordings instead of scores: “I want to hear what the thing sounds like, and there ain’t no way a sheet of paper sounds like Jimi Hendrix”. (Katz, 2010, p. 32)

There is a clear emphasis here on the importance of learning to attain a specific sound (i.e., a timbre).

That timbre was seen as central to the success of a performance can also be seen in people’s listening habits. Indeed, in recorded music it is often the sound that listeners single out as significantly impacting their overall musical experience. For example, in his 1976 review of the work of singer Janis Joplin, Jack Shadoian wrote:

> Kozmic Blues was bound to be a disappointment, and it was. Janis seemed displaced. The new band didn't help much and her voice, subjected to studio clarity, sounded more strained than expressive. Her style, too, transplanted to a tighter setting, seemed overblown and uncontrolled. (Shadoian, 1971, para. 6)

The disappointment of *Kozmic Blues* was not in its execution, or its musical composition, but in its sound. The strict studio setting in which the recording took place made Joplin's voice sound too "strained", "overblown and uncontrolled" (Shadoian, 1971, para. 6). Clearly, vocal timbre in this instance is a musical feature that is considered key to the success of this musical performance.

Another example of the importance of timbre to the success of a song can be seen in an interview by Dan Wilson on recording Adele's "Someone Like You". While discussing the process of recording the song, Wilson recalls that:

> Once we started recording, I was very much concentrated on making sure we got a killer vocal performance, because I was starting to think this was a special recording, and also she is such a pleasure to record! She sounds so great coming back out of the speakers, and I was dead set on making the song sound great but very natural, very vulnerable, very devastated. On the second day, her voice had a rougher, more ragged edge, and I suggested we go back and re-record the last chorus so it would sound more emotional. And it did, it was heartbreaking. (Waterman, 2012, para. 9 – 10).

It is telling that Wilson considered the recording taken on the second day as more successful not because of Adele improves in areas of pitch and rhythm, but because of the quality of her *vocal timbre*. Again, we see the success of the performance hinged more on the quality of the singer's voice, than on other technical aspects. This emphasis on timbre and vocal timbre that can be seen throughout many reviews of popular/recorded music demonstrates that timbre is considered an important and expressive musical feature.

This emphasis on timbre is further exemplified in other new forms of music in which the timbre of different sounds is a primary compositional tool. Electroacoustic music is one example of this. This music's creation typically relies completely on the interaction of, and relationship between, different sounds (see, e.g., the musical works of Dennis Smalley and Pierre Schaeffer, as well as their academic works (Smalley, 1986, 1994, 1997; Schaeffer, 1966)). As technological developments allow for the creation of newer and more and more novel timbres and sonic objects, composers, particularly of electronic dance music, are increasingly considering timbre as a primary compositional device (Hill, 2005).

2.5 The Role of Technology in Foregrounding Vocal Timbre

Throughout history, there has been a strong interest in timbre, its role in musical contexts and its potential to communicate meaningful information.

Popular music is another musical form to follow on from this long tradition of timbre and vocal timbre manipulation, exploring vocal timbre in a new way facilitated by technological developments of the last century. I argue that these technological developments have served to further foreground the importance of timbre and vocal timbre in three key ways: by allowing it to be preserved, by allowing it to be manipulated, and by allowing it to be isolated. This section will briefly explore each of these ideas in turn.

2.5.1 Preserving vocal timbre.

Developments in recording, playback devices, and other production effects and performance equipment have, for the first time in history, allowed vocal timbre and its nuances to be preserved. Vocal sounds that would ordinarily only be accessible at close or intimate distances, or which would only be able to be heard at a live performance, are now able to be heard "on demand" and in a variety of situations (from the speakers of a shopping centre, to hearing a song through headphones, to the amplification of the voice at a concert). This has meant that listeners can now engage with vocal timbre in a way that, until the 20th century, had not been previously possible. Consequently, for popular music specifically, and music more generally, vocal timbre is a musical feature that is increasingly contributing "significantly to both the immediate pleasures and conceptual meanings afforded by this music" (Heidemann, 2016, p. 1).

Recording and playback devices have also allowed specific characteristics of a vocal timbre to become distinctive units for musical expression. As will be explored in Chapter 3, listeners are very good at identifying and distinguishing between instruments based only on their timbres. The distinctiveness of these timbres is, at least in part, due to these instruments having different overall makeups and methods of sound production.

Similarly, vocal timbres may also be highly distinctive as no two human bodies are exactly alike. The production of vocal sound consists of three components: "(1) the system below the larynx, (2) the larynx and its sounding structures, and (3) the structures and the airways above the larynx" (Stevens, 200, p. 1). Basic biological differences can impact the size and shape of these structures (Titze, 1989), thus impacting one's overall vocal timbre. For example, "thickness of the vocal folds, differences in the shape of a person's palate, and the dynamic use of the vocal tract, give rise to differences in pronunciation, accent and other" idiosyncratic features of one's vocal timbre (Lavan et al., 2018, Introduction). Furthermore, a single human vocal tract can be manipulated to produce a multitude of different vocal timbres (Lavan et al., 2018). Consider the differences in commonly heard vocal variations such as laughter, whispering, shouting, and speaking. Therefore, in the same way as different bodies may produce distinctive and unique vocal timbres, the use of a single body in different ways may produce equally distinctive variations in vocal timbre.

Thanks to the development of recording and playback devices, this distinctiveness can now be captured and preserved in a way never previously possible. Popular music is one musical form which has taken advantage of these developments. Consequently, the distinctiveness of vocal timbres has become a cornerstone of this music. Allan F. Moore has argued that elements such as timbre are indeed strong signifiers in popular songs:

> Serious analysis of popular song (as opposed to analysis of its lyrics) is a recent phenomenon, and much of the early work consisted either of working from transcriptions (for example, Wilfrid Mellers in the 1970s) or of producing analytical diagrams on paper (for example, Walter Everett in the 1980s). Many commentators (Everett included) have remarked on the relative inadequacy of these approaches, some arguing that *what we might define musicologically as categories like texture and timbre signify much more strongly than they do in music normally encoded in scores* [emphasis added]. (Moore, 2010, pp. 257–258)

In this way, features such as timbral variety have become important syntactic aspects of popular music. The preservation of this feature, however, relies on recorded sound—for how else, for example, does one capture timbres which are "not normally encoded in score" (Moore, 2010, p. 258)? In this way, technology has allowed vocal timbre to be foregrounded in our musical experiences as it allows the staggering versatility of the voice to be preserved and, thus, to be used for musical expression.

One way in which these vocalisations may be effective tools for musical expression is due to their potential to bond the musical experience with the human experience. For example, a wailing vocal timbre is distinctive compared to a smooth one, however a wailing vocal timbre is also distinctive and potentially meaningful as we may attribute this sound to an experience in our everyday lives—a wail may indicate to us that a person is in pain. In this way, vocal timbre can go beyond “purely musical” sounds to represent that which it embodies and which it intends (D. K. Blake, 2012, p. 11). That recordings can capture and preserve these distinctive and, potentially, communicative vocal qualities allows such nuances of vocal timbre to become important musical parameters. This is particularly true of vocal timbre in popular music in general, and popular songs specifically.

2.5.2 Manipulating vocal timbre.

Technological developments have also given creators and listeners the power to engage with and manipulate vocal timbre in new ways. Vocal timbre is no longer only able to be heard in live performance, rather the voices of our favourite singers can now be heard on demand. Nor must vocal timbre adhere to a specific pedagogical approach for pragmatic reasons (e.g., training the voice in a specific way to enhance projection), but it can be more nuanced and conversational while still being foregrounded (through microphones) and preserved (through recordings). In this way, technological developments facilitate the manipulation and foregrounding of vocal timbre.

This is evident in popular vocal songs where sound processing techniques such as compression, reverberation, and equalisation allow the voice to be manipulated and altered for creative and musical effect. The desire to engage with vocal timbre in this way may suggest that it is recognised as an important and, potentially, meaningful musical parameter. Specifically, for popular vocal songs, processing effects have been shown to impact perception. This idea is explored by Serge Lacasse in his 2000 thesis. For example, Lacasse argued that the use of reverberation in a song was closely connected with a listener's idea of space:

> In most of rock music, besides enlarging the performer's sonic image, it seems that the virtual environment created by reverberation acts also as a kind of theatrical ramp, a fence delimiting the territory: listeners on one side, the artist on the other. When this limit is overstepped, when the absence of reverberation makes the two territories merge together, there is a feeling of invasion, of intrusion, or of extreme intimacy. (Lacasse, 2000, p. 240)

The application of reverberation may create a sense of distance, or bring the listener closer to the performer. In this way, reverberation is one example of how manipulating the sound of the voice can place it in the fore of the listening experience and potentially exaggerate its emotional message.

The exaggeration of a vocal timbre's emotional message may be further heightened by the decision to include/exclude certain aspects of a vocalisation. Creators have a high level of control over a recording. They may remove undesired aspects of the voice from a final product. However, it is not uncommon for sounds which may be considered nonmusical to be retained in a recording. One such example is breath. In editing it is possible to reduce the audibility of the breaths a singer takes. However, there are many examples where very audible breaths are left in the final recording. For example, The John Butler Trio's "Revolution" (Butler, 2010, track 1). The inclusion of such sounds in the recording may serve to foreground vocal timbre as a musical feature, and to heighten the emotional message conveyed in the vocal line (indeed, breaths may create a sense of intimacy—after all, without amplification, one can only hear a breath if one is close to the singer). It may also serve to reinforce the link between timbre and the human experience as, in everyday dialogue, one often hears not only tone of voice[6] and words, but other sounds too (including the intake of breath). In this way, the power to include/exclude certain vocal sounds in a recording can heighten the conversational aspect of the voice, making it more relatable, more emotional, and, therefore, potentially more foregrounded.

2.5.3 Analysing vocal timbre in isolation.

The technological developments discussed in 2.5.2 serve to emphasise vocal timbre as they allow for it to be preserved and manipulated. However,

[6] The term "tone of voice" here refers only to the delivery of spoken words. A more in-depth definition is given in section 3.3.1 below.

these developments also allow for vocal timbre to be studied in a more systematic way. One can now hear a vocal timbre on repeat, and, through the use of isolated vocal tracks, can even hear it in isolation. The developments in spectrographs also mean that timbre can be visualised, and patterns can be visually identified (although spectrographs do have limitations, as discussed in Chapter 4). This ability to repeat and isolate vocal timbre has made it a much more accessible musical feature to study, resulting in it being more readily included in musicological discussions. David Brackett's *Interpreting Popular Music* is one such instance of vocal timbre in musicological study. In this book, Brackett discusses the role of vocal timbre in song and uses spectrographs to support his musical analyses (1995).

Technological developments in the scientific field have also resulted in vocal timbre being increasingly studied. For example, Serge Lacasse combines both musical analysis and scientific lines of enquiry to better understand the role of vocal staging in musical perception (2000). Robert Cogan and Pozzi Escot also take a scientific approach, this time objectively analysing the acoustic features of a vocal timbre through systematic and scientific methods (Cogan, 1987; Cogan & Escot, 1976).

From the musicological examination of vocal timbre in isolation, to the scientific study of vocal timbre's acoustic features, technological developments are increasingly facilitating a wide array of approaches to vocal timbre analysis. The increase in explicit and systematic studies of vocal timbre demonstrates this musical feature's prevalence and importance in music in general, and in popular vocal song specifically. In other words, such analyses solidify vocal timbre’s status as an important musical feature; the increased discussion of vocal timbre is evidence of its importance.

2.6 Conclusion

From this broad overview and from the examples included in this chapter, one can see that timbre has been an important part of music making and musical experience. Technological developments have further highlighted and foregrounded vocal timbre as they allow this musical feature to be used in new ways and studied in a more robust and systematic manner. Popular music in general, and popular vocal songs specifically, are musical fields where both vocal timbre and technology manifest themselves. For these reasons, vocal timbre is foregrounded, making such music a good resource for further research into vocal timbre.

3 Conveying Meaning and Emotion Through Nonlinguistic Cues

3.1 Introduction

Imagine the shock of a clap of thunder, the adrenaline rush after hearing a shrill scream, smiling at the sound of a loved one's voice. Sounds have the potential to signify a thing (a storm, danger, a familiar person), and to be associated with certain emotions (shock, fear, safety). This chapter will explore how sound may convey, and potentially impact perception of, meaning and emotion in both everyday and musical contexts. Through this exploration, this chapter aims to show that the existing literature lends support to the hypothesis that emotion expressed in vocal timbre impacts emotional perception of sung words.

3.2 The Importance of Sound

By virtue of being associated with a listener's perception of a physical thing (such as an animal call representing a particular animal) or abstract concept (such as a scream representing fear), sound can come to be associated with certain meanings and therefore act as carriers of those meanings. This section will examine three lines of investigation relating to how sounds may be perceived as meaningful by listeners:

1. The way sound can be altered through the use of space.
2. The relationship between sound and place *as time*, and sound and place *as location*.
3. The impact of environmental sound on perception of meaning and emotion.

These cases show how sounds may acquire certain connotations, and the potential impact of these connotations on the perception of meaning and emotion.

3.2.1 The relationship between sound, space, and place.

The relationship between sound and space can be observed in everyday life. For instance, the experience of the echo of a loud voice in a lecture theatre, or muffled steps in a carpeted hallway. This relationship is not only passively experienced, it can also be actively engaged with. For example, a solo flautist may select a performance space based on how well sound projects in that environment, an acoustician may choose a certain ceiling height to encourage favourable acoustics in a music concert hall. Consciously observing how sound behaves in space, and using space to manipulate sound, suggests that this relationship carries significance.

The field of architectural acoustics is one example of how the relationship between sound and space is intentionally manipulated. This field, widely recognised to have been founded by Wallace Sabine in the early 20th century, is focused on the relationship between sound and space. Instances of architectural acoustics can be found in many contexts—for example, in an airport (to reduce excessive noise), or in a theatre (to increase clarity of spoken words).

When Sabine first investigated the relationship between sound and space, his goal was to improve the intelligibility of words in the Fogg Lecture Hall at Ohio State University (Sabine, 1922, p. 3). One obvious driving force in architectural acoustics seeking to manipulate the relationship between space and sound in this way is that of practicality. For example, there would not be much point attending a lecture at the Fogg Lecture Hall if one could not hear or understand what is being said.

Nevertheless, although it may not be overtly stated, this manipulation of sound through space may also show a concern for the *aesthetic* relationship between sound and space. One can find instances which illustrate that a relationship between sound and space was observed in some contexts throughout history. For example, some theories suggest achieving a desired resonance was a consideration when constructing certain chambers in Egyptian Pyramids (Malkowski, 2010). Another instance of the relationship between space and sound being considered is in open-air theatres. The Hellenistic theatre of Epidaurus, on the Greek Peloponnese, is one famous example. This open-air theatre is well known for its well-designed acoustic properties which allow “sound coming from the middle of the theater [to reach] the outer seats, apparently without too much loss of intensity” (Declercq & Dekeyser, 2007, p. 2011). The reason for these acoustic properties is thought to be that “the rows of stone benches at Epidaurus affect sound bouncing off them” such that “frequencies lower than 500 hertz are more damped than higher ones” (Ball, 2007, para. 9). This acoustic design helps to filter out background noise while projecting the performer’s voice. That the architecture of this space has been copied elsewhere in Greek and Roman theatres (Ball, 2007) demonstrates that space was intentionally used to manipulate sound.

Certainly, there were practical reasons for manipulating sound in open-air theatres—the performer’s voice needed to be amplified and background noise minimised so that the audience could understand the words. However, there is evidence to suggest that aesthetic aspects were also an important consideration. For example, Marcus Vitruvius Pollio in the 1st century B.C. states that:

> By the rules of mathematics and the method of music, they sought to make the voices from the stage rise more clearly and sweetly to the spectators’ ears. For just as organs which have bronze plates or horn sounding boards are brought to the clear sound of string instruments, so by the arrangement of theaters in accordance with the science of harmony, the ancients increased the power of the voice. (As cited in Declercq & Dekeyser, 2007, p. 2011)

References to the “sweetness” and “power” (as cited in Declercq & Dekeyser, 2007, p. 2011) of the voice suggest that it is not only the words, but also the sound of the voice that needed to be projected. That the clarity of both words and tone was sought out suggests that aesthetic considerations, as well as practical ones, influenced the design of such spaces.

Another example of the relationship between space and sound being engaged with throughout history can be found in the suggestion that prehistoric cave paintings tend to appear in cave chambers that are highly resonant (Lacasse, 2000, p. 33). In one study investigating this phenomenon, it was found that:

> some signs were rediscovered aurally: by advancing in total darkness through the cave and presuming that a sign would be found in a particularly resonant place, locating the latter, switching on a light, and indeed finding such a sign, even in a place unsuited for paintings. (Reznikoff, 1995, p. 547, as cited in Lacasse, 2000, p. 33)

It seems that, rather than practicality, some other benefit or pleasure (potentially related to the acoustic properties of these caves) was derived from engaging with this space. These examples show that a relationship (not only practical, but aesthetic too) between sound and space has been observed throughout history, and that “the staging of voice is neither new, nor specific to Western cultures” (Lacasse, 2000, p. 30).

Just as sound and space can be related, so too can sound and place. This may occur in two ways: place in the sense of time and place in the sense of location. One example of the relationship between sound and place in the sense of time can be found in the use of Christmas carols, which are usually only sung at a specific time of the year. When hearing these carols and associated sound (e.g., the jingling of bells), the listener may be reminded of Christmas. In this way, the sound of carols is attached to a sense of place in time via the experience and memories of the listener.

This idea may be extended to any number of sounds, both musical and non-musical, and may evoke a range of potentially emotional responses. That is, if certain sounds are associated with particular moments in time, hearing these sounds (even when they are out of context) may evoke the emotional connection associated with these moments. For example, hearing the call of crickets may evoke happy memories of summer holidays. When hearing these sounds out of context the connotations of a specific sound may still have the potential to impact our perception of emotion and meaning.

In addition to sound and time, a connection between sound and location can also be found. A 2015 study by Brown and Basset explored the idea that a piece of music could be composed for a location. In particular, they explored the relationship between the performance of improvised piano and the location of the Auckland Town Hall Concert Chamber (Brown & Bassett, 2015). Because space has the potential to impact sound, if a piece of music is created for a location, then the sound of that piece will not be the same in a different location. In this way, by actively considering and engaging with space in the compositional process, the composer is effectively using location as an instrument. Changing location then is in effect changing the instrumentation of the piece. The result is a performance that is made for, and attached to, a particular place.

3.2.2 Environmental sound.

Environmental sounds can transmit meaning in a variety of ways. The first, and perhaps most obvious, way that environmental sounds can transmit meaning is for the purpose of survival. For example, in a jungle, a growl may indicate that an angry predator, and therefore danger, is nearby. In this way, a connection may be developed between a growl and the sense of danger. Similarly, in a city the sound of a siren, which may be associated with a bomb raid in times of war, may convey a similar sense of danger. In this way, listeners may learn to associate certain sounds with particular meanings. Because these environmental sounds are nonlinguistic (the predator does not *say* that it is dangerous, but their growl can *tell* us), extrapolating meaning (through a process of learned associations) from the sounds alone may be considered an important tool for survival.

Another reason one may expect environmental sound to carry meaning is that nonlinguistic communication happens between other animals, and presumably happened in prelinguistic humans. If humans could communicate prior to language, it is reasonable to assume that this communication must have been based on vocalisations (such as grunts, growls, whistling) or gestures. These utterances would not only have expressed the intent of the "speaker", but must also have been perceived as intending something by the "listener". It is the duality of signals both intending and being understood to intend something that qualifies them as communication, for "communication does not begin when someone makes a sign, but when someone interprets another's behaviour as a sign" (Burling, 2000, p. 30). It is this perceived intent that is important to this thesis as it suggests that sound on its own not only transmits meaning, but these sounds can have *intended meaning* (e.g., a grunt meaning *yes* can sound very different to a grunt meaning *no*). By examining the emotive qualities of these sounds, we may be able to identify which signals allow us to perceive emotional valence in vocal utterances.

3.3 Conveying Meaning and Emotion Through Tone of Voice and Body Language

There are features of spoken communication that are not linguistic themselves (i.e., they are para-linguistic) which can also impact perception of emotion and meaning in spoken words. It has been observed that words alone often "lack the capacity to carry the whole weight of a conversation, as our verbal lexicons are extremely poor in comparison with the capacity of our minds for encoding and decoding an infinitely wider gamut of meanings to which at times we must refer as ineffable" (Poyatos, 1992, p. 50). In this context, para-linguistic cues can be used to emphasize a message, "deemphasize it or contradict it altogether" (Poyatos, 1992, p. 51). This section surveys three lines of enquiry exploring how para-linguistic features may impact a listener's perception of meaning and emotion:

1. The impact of emotional tone of voice on perception of emotion and meaning of spoken words;
2. The use of a form of parent–infant communication called *motherese*; and
3. The impact of facial expressions on the sound of the voice.

3.3.1 Emotional tone of voice.

Often in verbal communication, the way something is said is just as important as what is being said (consider how statements like "nice hat" can be imbued with sarcasm). In this way, tone of voice is important for optimal verbal communication (Wilson, 2011). Tone of voice is a source of para-linguistic information. Timbre has been identified as a primary quality of para-linguistic features of speech (see, for example, Poyatos, 1975, p. 291; Zarate, Xing, Woods & Poeppel, 2015). Para-linguistic features and Tone of voice provides listeners with information about a speaker's emotional and attitudinal state through variations in such factors as loudness, pitch, syllable duration, intensity and prosody[7] (Krestar & McLennan, 2013, p. 1793). This

[7] Prosody refers to the rhythm and tune of spoken words at the suprasegmental level.

can be seen in the way that emotional tone of voice impacts a listener's perception of spoken words, and, in particular, words' emotional content. For example, Nygaard and Lundersl found that "emotional tone of voice affects the processing of lexically ambiguous words by biasing the selection of word meaning" (Nygaard & Lundersl, 2002, p. 583). That is, when the meaning of a word is not necessarily clear, the speaker's emotional tone of voice appears to influence how listeners perceive its meaning.

In addition to altering perception of words, emotional tone of voice can also help us to understand the meaning of words. That is, words are more easily (and quickly) understood when the emotional meaning of words and tone of voice accord. A 2008 study by Nygaard and Quees examined this. It investigated how linguistic and nonlinguistic information was integrated in the processing of words with intrinsic emotional connotations (Nygaard & Quees, 2008, p. 1018). It was found that faster processing of spoken words was facilitated when emotional tone of voice was concurrent with the intrinsic emotional connotation of the word (e.g., a happy word spoken in a happy tone of voice). It was also found that the integration of para-linguistic and linguistic elements of speech occurred relatively early in the perception process, suggesting that the perception of spoken words is not independent of perception of emotional prosody (Nygaard & Quees, 2008, p. 1025). In summary, "[t]his congruence effect suggests that emotional prosody provides a specific semantic or communicative context that facilitates and shapes lexical processing and selection" (Nygaard & Quees, 2008, p. 1025).

Thus, emotional tone of voice can both impact our perception of, and help us to understand, words. This suggests that emotional tone of voice can affect semantic processing, perhaps by facilitating access to relevant lexical information. However, this effect on semantic processing takes time. A 2013 study by Krestar and McLennan investigated the impact of variations in tone of voice within a given speaker. In this study, the authors examined the effect of intratalker[8] variations in emotional tone of voice on spoken word comprehension. This study found that when the time–course of spoken word recognition and processing is relatively fast, then word processing was affected less by variation in emotional tone of voice. In these circumstances, therefore, only linguistic properties are considered when a word is processed for meaning, stripping away all nonlinguistic information such as emotional tone of voice, and arriving at what is called the abstract representation of a word (Krestar & McLennan, 2013, pp. 1793–1794). Conversely, when spoken-word recognition and processing was delayed, word processing was affected more by variations in emotional tone of voice. In this case, nonlinguistic features such as emotional tone of voice are preserved and utilised when a word is processed for meaning (Krestar & McLennan, 2013, pp. 1793–1794). It is thought that the listener is then relying on an episodic representation of the word, that is, a memory representation based on previous encounters with this word in real life. Although Krestar and McLennan's findings do not align completely with

[8] Intratalker variation refers to the variation within a given speaker in linguistic–structural effects including phonetic and prosodic features, discourse and lexical factors, and emotion (Wright, 2006, p. 1).

those of Nygaard and Quees (2008), both studies show that emotional tone of voice has the potential to impact perception of spoken words.

Tone of voice can also impact our perception of a person in a more general sense. In his book *Speech, Music, Sound,* van Leeuwen discusses how listeners may derive meaning from certain features of speech, in particular, meaning about relationships (van Leeuwen, 1999, p. 24). For example, speaking to someone in a soft tone of voice evokes more intimate connotations than speaking in a full, loud voice (van Leeuwen, 1999, p. 24). In this way, tone of voice may also convey, and be used to convey, meaning about the relationship between speaker and listener.

That tone of voice may impact a listener's perception of a speaker's personal qualities has also been explored by Lewis in a 2000 study examining perceived leadership abilities. Here, Lewis tested three emotional states: neutral, sad, and angry (Lewis, 2000, p. 223). It was found that emotional tone had a significant effect on perceived effectiveness of leadership, with leaders who display negative emotions being perceived as less effective (Lewis, 2000, pp. 230–231). Emotional tone of voice also appeared to impact on *follower affect* (how moved to action the listener felt) such that:

> expression of sadness seems to reduce arousal, while leader anger increases follower arousal. The affect-related findings indirectly suggest that a leader's anger may create motivation to work harder to improve the situation on the part of followers, while a leader's expressed sadness may lead to passive acceptance rather than effort to make things better. (Lewis, 2000, p. 231)

Emotional tone of voice can also impact perception of a speaker's own emotional state. That is, if a person speaks in a sad voice, they are more likely to feel sad. In their 2015 study, Aucouturier et al. "created a digital audio platform to covertly modify the emotional tone of participants' voices while they talked in the direction of happiness, sadness, or fear" (Aucouturiera et al., 2015, p. 948). Participants heard modified versions of their voice while speaking. These modifications made the participants voice more or less happy, sad, or fearful. Participants were not aware that their voices were being modified and they were not consciously monitoring their own emotional signals. Interestingly, participant's emotional sates changed to more accurately reflect the emotion conveyed by their altered voices. In other words, hearing a happier version of their voice caused participants to feel happier. This finding suggests that we "use the same inferential strategies to understand ourselves as we understand others" (Aucouturiera et al., 2015, p. 948).

In sum, emotional tone of voice can affect perception of meaning of spoken words at a lexical and (inter)personal level. It can affect how we perceive meaning in spoken words by biasing selection of word meaning, and by making it easier and faster for us to understand word meaning. It also affects how we perceive emotion and personal attributes in others, and even how we perceive ourselves. Clearly, then, emotional tone of voice is a highly evocative and meaningful source of information. As will be explored in later sections (see section 3.5), emotional tone of voice may be likened to vocal timbre in that it is a feature which can accompany words but is itself not linguistic—i.e., it is a para-linguistic feature. They are also similar in consisting of acoustic properties of the sound signal. If emotional tone of voice can have such a strong impact on perception of words and of personal and interpersonal emotional states, then it is possible that vocal timbre too will have such an impact. Perhaps more so, given that vocal timbre occurs in the highly stylised context of song, in which emotive aspects of vocal timbre can be exaggerated for artistic effect.

3.3.2 Baby talk: infant–adult communication.

Given that attention to tone of voice seems to impact our perception in such a robust manner, and in such a number of everyday contexts, it seems likely that processing tone of voice for meaning is something that would begin to take place early on in our development. Existing research seems to confirm this is the case. Infants seem to be able to process aspects of the human voice (including emotion) quite well from an early age. This was shown, for example, in a 2010 study by Grossmann, Oberecker, Koch, and Friederici looking at the processing of voice specificity (the ability to discriminate voice from other auditory stimuli) and prosody specificity (the ability to discriminate between emotions in voices) in young infants (Grossmann et al., 2010). In the case of voice specificity, the authors found that 7-month-olds process voice in a “fairly specialized brain region” (Grossmann et al., 2010, p. 855), indicating an early developing specialised neural mechanism. However, since the authors found a difference between 7- and 4-month-olds in voice processing, this mechanism appears to develop with age (seemingly quite quickly) rather than being in place at birth.

Similar results were found in the case of prosody. Here, the authors found that 7-month-olds responded in adult-like ways to happy and angry stimuli. Grossmann et al. note that "this result indicates that the enhancement of sensory processing by emotional signals is a fundamental and early developing neural mechanism engaged to prioritize the processing of significant stimuli" (Grossmann et al., 2010, p. 855). In sum, it appears that infants develop the ability to distinguish between prosodies, and therefore to attend to para-linguistic cues such as tone of voice, very early on in their development.

Interestingly, it was also found that angry stimuli elicited a stronger response than did happy stimuli, indicating that "threatening signals have a particularly strong impact on voice processing" (Grossmann et al., 2010, p. 855). This finding is in line with adult-like behaviour, suggesting a "developmental continuity in how the human brain processes happy prosody" (Grossmann et al., 2010, p. 856). That 7-month-olds appear to process prosody in an adult-like way suggests that mechanisms for processing emotional signals in speech develop early in life and continue to be important for the perception of meaning and emotion.

The early sensitivity to these para-linguistic cues is exploited by parents through the use of a special form of infant-directed speech called *motherese*. Motherese (or baby talk) is the special speech register used by adults to address babies and is typically characterised by heightened fundamental frequency, exaggerated intonation contours, and hyper-articulated vowels (Burnham, Kitamura, & Vollmer-Conna, 2002, p. 1435). It is thought to assist with language acquisition, perhaps due to the augmenting of phonic characteristics of language (Burnham et al., 2002) and/or by engaging infants' attention (Fernald & Kuhl, 1987; N. A. Smith & Trainor, 2008).

In their 2002 study, Burnham et al. investigated the use of infant directed speech (i.e., motherese) by looking at the speech patterns of parents when speaking to infants, to pets, and to other adults. They found that, while there were similarities between infant- and pet-directed speech (heightened pitch and effect), infant-directed speech was unique in that it contained hyper-articulated vowels (Burnham et al., 2002). In addition, it was also found that "affect was greater in infant- than in pet-directed speech . . ., but affect in both infant- . . . and pet-directed speech . . . was higher than in adult-directed speech" (Burnham et al., 2002, p. 1435). In other words, when speaking to both pets and infants (who have little or no language), speakers used much more feeling and emotion than when speaking with adults (who have full command of language). While hyper-articulating vowels and exaggerating affect may improve language acquisition in infants, the fact that affect was highly exaggerated in both infant- and pet-directed speech may suggest that, in the absence of a linguistic form of communication, a speaker may rely on tone of voice to transmit meaning.

Exaggerating affect appears effective in maintaining an infant's attention. However, there are other aspects of motherese that seem to have an attention-grabbing effect. One such aspect is the many frequency modulations that occur in motherese speech. Interestingly, these modulations impact the tone colour of the spoken voice (i.e., the timbre of the voice). It is possible that both affect and frequency modulation (and by extension timbre) are linked: one cannot express affect in a monotone voice; and frequency modulation may convey affect as pitch (and by extension tone of voice) may come to be associated with emotional expression (e.g., a high, bright voice meaning happiness).

A 2017 study by Piazza, Cătălin Iordan, and Lew-Williams found that parents do in fact alter the timbre of their voices specifically when speaking to infants. These timbre alterations seem to occur in a similar way across language groups, perhaps suggesting that there is some universality to the way timbre is used for expression and communication (Piazza, Cătălin Iordan, & Lew-Williams, 2017). If it is that timbre variations are used in a consistent manner, then, as suggested in sections 3.2 and 3.3.1, these variations may also come to be associated with different emotional messages.

It is generally acknowledged that infants are more attentive to spoken words when they are delivered through motherese speech patterns (Fernald, 1985, p. 181). In their 1987 study, Fernald and Kuhl investigated which attributes contribute to this “attention grabbing” nature of motherese. It was found that infants exhibited a preference for speech that featured pitch variations, but not for speech that featured variations in rhythm or dynamics (Fernald & Kuhl, 1987). That is, “it is the increase in F0-modulation… [pitch modulation] within an expanded range, rather than the higher absolute F0-level, that makes motherese speech more interesting to infants than adult-directed speech" (Fernald & Kuhl, 1987, p. 290). In other words, it is the variation in fundamental frequency in motherese speech that is attention grabbing; simply speaking in a high pitch voice does not have the same effect. Interestingly, this change is also more likely to impact the sound (timbre) of the voice more than other elements such as rhythm and volume.

Indeed, timbre is an element that infants have shown the ability to discriminate. In their 1988 study, Clarkson, Clifton, and Perris found that infants show the ability to "analyze the spectra of tonal complexes and discriminate differences in one of the most important cues for timbre perception in adults, the spectral envelope"[9] (Clarkson et al., 1988, p. 15).

That a similar result was found in this research suggests that processing of timbre also begins at an early age (i.e., from infancy).

[9] According to Clarkson et al.:

> "[t]he spectral envelope is defined as the curve connecting the points that represent the amplitudes of stimulus components in a tonal complex. The shape of the spectral envelope can be changed in a variety of ways by manipulating the amplitudes of stimulus components and by varying the frequencies of those components." (Clarkson et al., 1988, p. 15)

In sum, motherese speech seems to be effective in drawing infants' attention to the voice by exaggerating certain, attention grabbing, qualities such as affect and varying speech frequencies. The above research has shown that infants develop the ability to attend to such qualities early on. Therefore, it may be that variations to the timbre of the voice achieved through motherese result in associations between tone of voice (timbre) and affect to be made from a very young age.

3.3.3 Facial expressions.

Facial expression is a form of nonverbal communication that is part of a system of body language (Reevy, Ozer, & Ito, 2010, p. 137). Humans are well disposed to understand and interpret emotional signals and meaning from such body language, and to use body language in conjunction with linguistic and para-linguistic cues to derive information (Reevy et al., 2010, pp. 137–139). In all cultures facial expressions have been found to be a reliable means of conveying emotions (Ekman, 1994; Ekman & Friesen, 1971; Kayser, 2016). For example, masks used in tragedy and comedy performances in ancient Greece and Rome served to exaggerate facial expressions (Kayser, 2016, pp. 21–22). The features in these masks were not arbitrary, but it was understood that they depicted certain fundamental emotional states. This is shown through such thinkers as Aristotle who wrote that "[t]here are characteristic facial expressions which are observed to accompany anger, fear, erotic excitement, and all other passions" (as cited in Russell & Fernández-Dols, 1997, p. 3).

As observed above, certain facial expressions may come to be associated with certain feelings and emotions. One may assume certain facial expressions when expressing particular emotions and feelings. In this way, an association may be made between an emotion/feeling, a facial expression, and other cues such as tone of voice. These associations may be drawn on in the *perception* of emotion too. Indeed, research has found that listeners have a well-developed capability for identifying facial expressions based only on vocal sounds. One example is "hearing" a smile. Several studies have shown that listeners are able to identify the facial expression of a smile based on the sound of vocalisations (Tartter, 1980; Tartter & Braun, 1994; Zacher & Niemitz, 2003). In one study, it was observed that "particular cue combinations appear to be heard as smiling specifically, whereas others are associated with emotionality in general" (Tartter, 1980, p. 24).

According to the Memetic Hypothesis, the ability of one to "hear a smile" suggests that our physical experience of expressing emotion may play some part in our perception of that emotion in others

> …by way of a kind of physical empathy that involves imagining making the sounds we are listening to. This is a special case of the general human proclivity to understand one another via imitation, which we can refer to as mimetic cognition or mimetic comprehension… (Cox, 2011, para. 2 – 3).

In other words, when hearing a certain tone of voice, a listener may reproduce within themselves the feeling of producing that tone. This process of imitation may contribute to how a listener perceives emotion. In his 2011 paper, Cox suggests that the memetic hypothesis may also explain how one perceives music as it

> …addresses the matter of embodiment by showing how musical imagery—recalling, planning, or otherwise thinking about music—is partly motor imagery. Motor imagery is imagery related to the exertions and movements of our skeletal-motor system, and in the case of music this involves the various exertions enacted in musical performance. The mimetic hypothesis details how this might play out and suggests how it might underlie conceptualization and meaning. (Cox, 2011, para. 2 – 3).

In this way, the system for understanding emotion in others through a process of imitation may also extend to music.

3.4 Music Conveying Meaning and Emotion

3.4.1 The processing of musical features.

In the above two sections, it has been shown that para-linguistic aspects of speech can convey meaning and emotion. The same may be said of music. To see this, one only has to turn on the television. In a movie, we know a scene is suspenseful because of the high tremolo in the strings. When watching the news, we know a story is just breaking because of the militaristic, heralding nature of the trumpets playing in unison. Specific musical elements have been found to convey meaning and impact emotion in a way that is (almost always) quite consistent across listeners.

Harmony is one such musical element. Harmony has been found to set up an expectation in music which can later either be confirmed or violated (Tillmann & Bigand, 2002). This is a form of *priming*, a term referring to the influence exerted by a given stimulus (a prime) on a person's response to a subsequent stimulus (a target). In their 2002 review of priming experiments, Tillmann and Bigand found that harmonic priming can occur in both local and global contexts. The local context refers to one prime chord or word followed by one target chord or word. An example of local context priming in language may be the target word *nurse* being more readily understood when it followed the prime word *doctor* compared to when it followed the prime word *lawyer* (Fromkin, Rodman, & Hyams, 2011; Tillmann & Bigand, 2002). This effect in language is called *semantic priming*. Similar effects are also generated by harmonic primes, with a target chord being processed faster when it is closely harmonically related to the prime chord, rather than when it is distantly harmonically related (Tillmann & Bigand, 2002). For example, a six-chord harmonic progression in C major (the *prime*) sets up an expectation for a cadence also in C major (the *target*). If the progression cadences instead in D major, then the prime and the target are incongruent, the expectation is most likely violated, and a processing delay is observed.

The global context, on the other hand, refers to a series of chords (as in a progression) or words (as in a sentence) acting as a prime, and finishing with a target word or chord. This context tends to be more ecologically valid, that is, closer to the "real life" experience. In language, research has shown that "the processing of a target word was facilitated if that word formed a congruent ending for the sentence context . . . [compared to] . . . when it formed an incongruent completion" (Tillmann & Bigand, 2002, p. 233). For example, the target word *pistol* is facilitated by the prime *the cowboy fired the* rather than by the prime *the interpreter knew the* (Tillmann & Bigand, 2002, p. 233). A similar effect was found for harmonic priming with participants' intonation judgments tending to be faster and more accurately when the final chord was expected. This effect of priming in harmonic progressions appears to be like that in language, suggesting that music can act and does act as a prime, in this case priming the listener for expected chordal endings. While this in itself does not necessarily imply that harmony carries meaning in the usual sense of the word, it does point to the existence of intrinsic relationships within this musical element. These relationships seem to be similar to those known to exist between words due to shared semantic features. In summary, harmony seems to act as a strong prime, impacting our perception in a robust and intersubjective way. This may lend credence to the idea that other musical elements, such as vocal timbre, could also facilitate priming in this way.

The similarity between harmonic priming and linguistic priming was more clearly shown in a 2004 study by Koelsch et al. In Koelsch et al.'s study, participants were played a sentence or musical excerpt (the prime) recorded from commercially available CDs and were then presented with a list of target words. The target words consisted of 44 German nouns (e.g., *wideness, narrowness, needle, cellar, stairs, river, king, illusion*) (Koelsch et al., 2004, p. 302). The target words were either related or unrelated to the linguistic or musical prime. Target words were related to musical primes in two ways: by composer self-report and by musicological terminology, as explained in Koelsch et al.:

> One-third of the musical primes . . . had been chosen based on self-reports of the composers. For example, the prime for the word needle was a passage of Schönberg's String Terzett in which he described stitches during his heart attack. The other musical primes had been chosen based on musicological terminology. For example, the prime for the word narrowness was an excerpt in which intervals are set in closed position (covering a narrow pitch range in tonal space, and being dissonant). . . .
>
> Most of the musical stimuli that primed concrete words resembled sounds of objects (e.g., bird) or resembled qualities of objects (e.g., low tones associated with basement, or ascending pitch steps associated with staircase). Some musical stimuli (especially those used as primes for abstract words) resembled prosodic, and possibly gestural cues that can be associated with particular words (e.g., sigh, consolation). Other stimuli represented stereotypic musical forms or styles that are commonly (that is, even by nonmusicians) associated with particular words (e.g., a church anthem and the word devotion . . .). (2004, p. 303)

The authors found that the “priming effect did not differ between language and music with respect to time course, strength or neural generators” (Koelsch et al., 2004, p. 302). In other words, this study suggests that lexical processing can be facilitated by both musical and linguistic primes, and supports the idea that harmony can and does convey meaning.

Further relations between linguistic and harmonic processes have been found, showing that the relationship between processing of musical and linguistic elements is multidirectional. For example, a study by Poulin-Charronnat, Bigand, Madurell, and Peereman (2005) that addressed the effect of harmony (as accompaniment) on vocal music perception found that music accompaniment can affect the distribution of a listener’s attention, resulting in semantic priming in language becoming secondary to harmonic priming (Poulin-Charronnat et al., 2005, para. 1). This relationship between semantic and harmonic priming is important for the present thesis, which examines the relationship between vocal timbre and lyrics. This research suggests that musical and linguistic elements interact in a complex way, and that musical elements are not necessarily subordinate to linguistic ones.

It is worth noting here the possibility for priming to occur at a sensory level only, which would negate the conclusion that harmonic priming is targeting lexical meaning. Sensory priming operates in such a way that “a chord sharing component tones, or overtones, with a preceding chord will be more highly anticipated than a continuation containing no overlapping frequencies with its predecessor” (Schmuckler as cited in Bigand, Tillmann, Poulin-Charronnat, & Manderlier, 2005, p. 1350). If sensory priming were to underpin faster recognition of target chords, then in global contexts harmonic priming should only occur in the final two chords of a progression. This is because the final two chords in a progression form a cadence. For example, chord V, to chord I. Cadences are highly characteristic and the shared overtones of chord V to chord I could prime the listener on a sensory level to expect chord I. However, this is not the case. Studies have found that, for harmonic progressions containing eight-chord sequences, the expectation of the target chord was not significantly influenced by the cadence, even when the cadence was harmonically correct (Tillmann & Bigand, 2002, p. 233). Instead, it was found that the six chords preceding the cadence more readily facilitated perception of the target chord. It has also been found that harmonic priming in a series of chords is stronger when the target is the tonic rather than a less related chord, even if the less related chord has occurred more often and shares more overtones with the prime (Bigand et al., 2005). If harmonic priming were only the result of sensory level priming, it is unlikely that these phenomena would be observed.

While harmony is a musical feature which can impact perception of meaning, various studies have shown that music in general also has the potential to impact listeners in a variety of ways. For example, it has been found that musical cues may be interpreted in a meaningful way, creating within a listener a strong and robust sense of expectation (Huron, 2007). Music has also been found to modulate attention in general (Anderson & Fuller, 2010), and, most critically, emotion specifically (Behne, 1997), and to increase attentional resources through stimulating emotion (Soto et al., 2009). While the nature of this relationship between music and emotion is not yet well understood (Juslin & Sloboda, 2001), it is nonetheless a robust one that permeates throughout our listening experiences.

In summary, research shows music has the ability to impact perception of meaning and emotion. Harmonic priming specifically has been found to be a robust process that is not subordinate to linguistic priming, but rather functions in conjunction with it. The priming paradigm used in the studies of harmony above may also be applied to other musical features such as (in the case of this thesis) vocal timbre. Music has also been found to modulate attention and emotion in a variety of settings. This is again useful to this thesis which queries wither vocal timbre can impact the emotional perception of lyrics. Although specific research into *vocal timbre* in this regard is yet to be done, as a salient musical element itself, it seems likely that it will also have the capacity to impact perception of meaning and emotion.

3.4.2 The processing of timbre.

Like harmony, timbre has the potential to affect how music as a whole is processed by the listener. This is evident in two ways. First, timbre can impact the perception of pitch and intonation. Second, timbre provides information about instrument identity. Such research suggests that timbre, rather than being a simple sensory phenomenon, adds another layer of meaning to musical elements. Although the specific nature of vocal timbre‘s impact on perception is largely underexplored, if timbre is shown to impact perception of other musical elements (such as pitch and instrument identity), then it is not unreasonable to expect this will also occur for lyrics.

3.4.2.1 Timbre and pitch perception.

Timbre has been found to impact one's perception of pitch. In their 2013 study, Zarate, Ritson, and Poeppel (2013) explored the impact of timbre on interval discrimination. Here, the authors tested interval discrimination across four timbres: voice, piano, flute and pure tones. It was found that changing timbres within an interval can "significantly affect interval discrimination" (Zarate et al., 2013, p. 8). In particular, it was found that "the varied spectral energy of instrumental timbre can alter pitch perception and/or interval discrimination." (Zarate et al., 2013, p. 8), resulting in some instrumental timbres facilitating better interval discrimination than others. For example, participants were more accurate in discriminating amongst intervals played with piano timbres rather than flute timbres, perhaps because of the onsets in piano tones being sharper than those of flute tones.

A similar effect can be observed, although not as strongly, in more ecologically valid contexts (such as in the context of a whole melody). A 2002 study by Warrier and Zatorre investigating the impact of timbral changes in such contexts found that timbre did influence pitch perception; however, this influence decreased as tonal context increased (Warrier & Zatorre, 2002, p. 206). One explanation for this is that there is enough pitch information in a whole melody context to set up a tonal framework, meaning the listener is less likely to be influenced by changes in spectral phenomena such as timbre.

Nonetheless, this research shows that timbre has the potential to impact pitch perception to varying degrees in musical contexts.

Timbre also has the potential to impact perception of intonation, i.e., how in tune an instrument is playing. In their 2015 study, Geringer, MacLeod, Madsen, and Napoles investigated the impact of vibrato, which can impact an instrument's timbre, on perception of intonation within instrument groups. Across all instruments tested (voice, trumpet, and violin), melodic intervals played without vibrato were perceived as more out of tune than those with vibrato. The researchers also found that:

> melodies performed with vibrato were judged differently: Violin was judged as least in tune for intervals mistuned in the flat direction, trumpet was heard as least in tune for intervals mistuned sharp, and voice was judged least in tune when intervals were in tune (relative to equal temperament) (Geringer et al., 2015, p. 675).

That the addition of vibrato, which has the potential to alter an instrument's timbre, resulted in different instruments being perceived as more or less in tune underscores the potential of timbre to impact perception of other musical elements—in this case, intonation.

3.4.2.2 Timbre and instrument identity.

Timbre has been found to impact perception about instrument identity. For example, most people can distinguish between a violin and a flute based only on their sound, even if these instruments are playing the same pitch at the same volume. This is because every instrument has a characteristic sound (i.e., timbre). Two key aspects of an instrument's timbre appear to play an important role in the perception of its identity: onset and pseudosteady state. Onset has been identified as a highly characteristic aspect of an instrument's timbre. Indeed, it has been found that participants can accurately identify an instrument based only on a recording of its onset (Saldanha and Coroso, as cited in Erickson, 1975, p. 61). Fourier analysis provides clues as to why onset may be so characteristic. Fourier analysis is a method of calculating frequencies (harmonics) based on a given fundamental. In music, a given fundamental generates a unique series of harmonics. While manipulating the speed of the onset can result in a greater or lesser spread of harmonics, each fundamental will always occur with the same unique harmonics (i.e., a flute playing middle C will always generate the same harmonics regardless of whether the note is played short and sharp, or smoothly). Thus the series of harmonics present in the onset may be important for instrument identification.

The second aspect of sound that appears to be important for instrument identification is the pseudosteady state. After the onset, sounds that have a long enough duration (such as a bowed note on the violin) enter the pseudosteady state. In this state, the tone must then be kept "alive" through timbral nuances (Cogan & Escot, 1976, p. 69). Vibrato is one way to create such timbral nuances, sustaining a tone by generating harmonics that constantly change. Spectral glide, the "modification of the vowel quality of a tone” (Erickson, 1975, p. 72), is another way of prolonging a tone. Keeping the sound alive through the pseudosteady state invariably means manipulating the instrument’s timbre (e.g., through vibrato or spectral glide). These manipulations may result in timbre being placed in the forefront of the listener’s attention.

While the research cited so far in this section focuses on instrumental timbre, vocal timbre would probably also elicit similar results. The sound of the voice is not a simple, single-faceted object, but a rich and complex one (Bonada & Loscos, 2003; Rodet, Potard, & Barrière, 1984). It is unique and it contains many characteristic complexities which are difficult to synthesise mechanically. Critically, vocal timbre contains equally characteristic and unique Timbral Attributes, both characterising the instrument of the voice and that of the individual singer. The difference between the sound of Janis Joplin’s vocal timbre compared to Joni Mitchell’s vocal timbre is almost as easy to notice as the difference between the sound of the sung voice and the sound of a trumpet, for example. This unique sound of the vocal timbres of different singers may be related to the unique sound of people’s spoken voices. Consider, for instance, one’s ability to instantly pick that the disembodied voice on the other end of the phone belongs to one’s best friend. If nothing else, this points to the capacity of timbre in general and probably vocal timbre in particular to convey information.

3.5 The Similarities Between Vocal and Instrumental Timbre, and Vocal Timbre and Emotional Tone of Voice

In the previous sections it has been shown that listeners perceive emotion and meaning in para-linguistic features, and the same can be said of certain musical features. Timbre in particular has the potential to impact perception of other musical elements, and can convey meaning on its own (e.g., about pitch, intonation, and instrument identity). The hypothesis of this thesis, i.e., that emotion conveyed through vocal timbre can impact a listener's perception of meaning and emotion in sung words, rests on the assumption that vocal timbre shares critical features with instrumental timbre and emotional tone of voice and, thus, can also be expected to affect perception of meaning and emotion. This will be argued in sections 3.5.1 and 3.5.2.

3.5.1 Vocal timbre and instrumental timbre.

Timbre is a highly salient musical parameter. Its unique characteristics impact perception of other musical elements, provide us with information about its source (i.e., instrument), and its manipulation is essential for maintaining listeners' interest. Although research on its relation to the issue of meaning and emotion does not examine vocal timbre in particular, insofar as instrumental timbre is just an instantiation of musical timbre more generally, as is vocal timbre, the findings of the research on instrumental timbre should apply to vocal timbre as well.

Voice is a highly distinctive auditory cue in both musical and real-life contexts (see, for example, Lavan et al., 2018; Stevens, 2000; Titze, 1989, as well as sections 3.3.1, and 3.3.2). On a very basic level, this distinctiveness may be attributed to the voice being in and of the body. Every human body is different, and as such, every vocal tract and resonating shape is different. Therefore, each voice will have subtle differences too. This makes vocal timbre unique among musical elements as no two vocal instruments are the same.

Nevertheless, as a type of timbre, vocal timbre should be expected to show similar effects to other instrumental timbres. An instrument's timbre is characterised by such things as how the sound is produced (blown, bowed, struck, etc.) and how the instrument resonates (e.g., the double reeds of an oboe vibrating together and resonating through the body). In the studies reviewed in previous sections, although many different types of timbres were tested (a flute is blown into, and the airstream resonates in the body; a violin is bowed, and the vibrating string resonated through the instrument), all were found to be equally capable of impacting perception of pitch and intonation. Vocal timbre may be more capable considering the immense potential for timbral flexibility. For example, opera singing and everyday speech produce strikingly difference in vocal qualities. Yet, there is nothing particularly unique about the opera singer's vocal tract except for the way the singer uses it (Sundberg, 1977, p. 91). A singer may also produce a number of distinctive vocal timbre which are not able to be produced on conventional instruments, but which nonetheless communicate meaning to a listener. For example, "the range of vocal techniques and timbres employed by Annie Lennox, Kate Bush or Bjork … [show] a capacity to portray heterogeneous voice qualities" (Middleton, 2010, p. 28). Some voice qualities which have been shown to convey meaning are cry breaks (Huron, 2010) and twang (Sundberg & Thalen, 2010).

Extrapolating from the findings on timbre more generally, vocal timbre could then affect processing of popular vocal song in the following ways:

1. In terms of pitch and interval perception.
2. In terms of singer identity (an extension of instrument identity).

Additionally, I propose that onset may be similarly important for the identification of emotional valence in vocal timbres. That onset is important in the identification of instruments based only on their timbre (see section 3.4.2.2) and may impact interval discrimination (see section 3.4.2.1) suggests that onsets have the potential to impact listeners' musical experiences in a robust way. Emotions may impact the sound of vocal onsets as certain emotive states can affect vocal production by altering the shape of the body, air flow and so on (as touched on in the introduction and as discussed in more detail in section 3.5.2). Given that listeners tend to draw information from onsets in musical contexts, and given that emotion may impact the sound of an onset, it may be that certain onsets become associated with certain emotive states. For example, if one associates a very breathy, weak, vocalisation in a spoken-word contexts with sadness, then each time similar acoustic features are heard in the sung voice, they may carry a similar connotation of sadness.

In 3.4.2.2, it has also been shown that timbre may be used to create and maintain interest. In the case of instrumental timbre, this may be achieved through variations such as vibrato during the pseudosteady state. Vibrato may also be added in the sung voice. However, because the voice has the potential to include a large number of timbral variations, it may be that a number of different variations are added to the sung voice in the pseudosteady state to maintain interest (e.g., tension, roughness, breath). These variations do not only exist in the musical context, but are often present in everyday situations—consider a voice shaking in fear, or bright and strong with confidence. In this way, these variations may not only help maintain interest, but may also convey and impact emotion due to their intrinsic associations with particular emotional states. If such variations impact the spoken voice in everyday contexts, then they are likely to also play a role in song.

3.5.2 Vocal timbre and emotional tone of voice.

The hypothesis presented in this thesis, that emotion conveyed through vocal timbre impacts perception of meaning and emotion in sung words, is particularly grounded in the idea that there exist similarities between vocal timbre and emotional tone of voice. That there is a relationship between tone of voice and vocal timbre seems likely, given that both are produced by the same instrument and both are used when expressing linguistic content.

Tone of voice is one way meaning and emotion are highlighted in speech, as evidenced by the priming studies discussed in 3.3.1 (Aucouturiera et al., 2015; Krestar & McLennan, 2013; Lewis, 2000; Nygaard & Lundersl, 2002; Nygaard & Quees, 2008; van Leeuwen, 1999). It is likely that a parallel mechanism is also available to achieve equivalent effects in song, more specifically, in popular vocal song. This is because popular vocal songs tend to convey emotion and meaning through the vocal line. Consequently, it is important that word meaning and emotion are readily comprehensible in the same way that the word meaning and emotion are readily comprehensible when communicating in spoken words. Therefore, it is likely that vocal timbre and lyrics are employed by performers and drawn on by listeners in the expression and perception of emotion and meaning in a similar, yet exaggerated, way to their spoken word counterparts in order to maintain clear communication and maximise emotive impact.

Further, both tone of voice and vocal timbre are para-linguistic features that are impacted by the physical effects that emotion has on body shapes and, more specifically, on the shape and functioning of the vocal tract. In the case of tone of voice, this has the potential to be introduced unconsciously by the speaker and, in the case of vocal timbre, a stylised version of this "natural" response has the potential to be included in song. In this way, this feature, which is a natural and unintended consequence of the speaker's emotional state, is likely to also manifest itself in song.

Research has shown emotional tone of voice can bias selection of word meaning (Nygaard & Lundersl, 2002; Nygaard & Quees, 2008), that the sound of the voice can impact perception of emotion and meaning (Krestar & McLennan, 2013; Nygaard & Lundersl, 2002; Nygaard & Quees, 2008), and that a listener's own emotional tone of voice has the potential to impact their perception of their own emotive state (Aucouturiera et al., 2015). This suggests that heard emotional signals must influence our perception of emotion and meaning in a robust way, especially if they are to compete with, and alter, a listener's internal emotional perception. If these emotional signals are so important in emotional tone of voice of spoken words, then it is also possible that they impact perception of emotion in vocal timbre in a similar way.

It is also likely that auditory links may be formed between body language and vocalisations. Given that similar mechanisms are employed to produce both spoken and sung words—the body as a resonating chamber, the use of the vocal tract to produce words—and given that body language seems to be the same in spoken and sung words—if we feel happy, when we speak we smile; if we feel happy, when we sing we smile—then the same auditory connections between body language and spoken words may be present for sung words.

What's more, these "emotional signals, either visual or auditory, can be considered as aspects of both an emotional response and social communication" (Adolphs, 2002, p. 169). That is, these signals can both be the result of feeling an emotion (I smile when I am happy) and be used as a means of communicating an emotion (smile signifies happiness). In this way, emotional tone of voice and body language both convey and are perceived as conveying emotion in everyday contexts. Viewing singing as a stylised form of speech, in which emotional signals are ingrained, it is possible to see how listeners may extrapolate emotion and meaning from vocal timbre in popular vocal songs.

Applying theoretical tools from linguistics to the analysis of timbre and lyric delivery has been shown to be a fruitful avenue of research (see, for example, Bauer, 2008; Huang & Huang, 2008). Specifically, viewing the sung voice as "a stylised means of conveying emotion using, among other things, paralinguistic features borrowed from everyday speech" (Lacasse, 2010b, p. 142) is becoming increasingly common. Certain genres of popular vocal song have been identified as drawing heavily on this connection. For example, there exists in rock singing a "'naturalistic' tendency" in which "… para-linguistic dimensions [are] often as important as direct verbal meaning" (Middleton, 2000, p. 29).

It has also been observed that the relationship between a singer's spoken and sung voice may be such that "the paralinguistic features used by singers become part of their singing style" (Lacasse, 2010b, p. 142). Several levels on which this connection can be made are identified by Serge Lcassse. These include:

- the generic level, where a singer's accent impacts vocal quality,
- the performance level, where a singer's unique speech patterns impact vocal quality, and
- the character level, where a singer may "act" out a performance of a song (Lacasse, 2010b, p. 142).

Lacasse also observes that a connection between the spoken and sung voice may occur on the microacoustic level (Lacasse, 2010b, p. 143). It is at this level that the form of certain acoustic features play a key role in conveying emotion. While all of the four levels outlined above are relevant to this thesis, it is the microacoustic level that is most pertinent as it is on the microacoustic level that the Vocal Timbre Features discussed in section 7.3.1 are able to be assessed.

3.6 Conclusion

The research explored in this chapter has shown that the communicative nature of sounds alone has been observed throughout history, and, therefore, that sound on its own is important. Evidence of this ranges from the locations of prehistoric cave paintings, to the Boston Symphony Hall, to the crafting or selecting of spaces based on the aesthetics of acoustics. Listeners can also perceive meaning and emotion in environmental sounds by means of learned associations that can occur between a sound (such as a growl) and the real-life implications of that sound (danger).

Para-linguistic and nonlinguistic cues also convey, and impact on perception of, meaning and emotion in everyday contexts. This appears to begin from an early age with the nonverbal communication between parent and infant, and continues throughout life in the perception of emotional tone of voice. The fact that emotional tone of voice has been shown to impact perception of spoken words in a robust way suggests that it is important for communication and that it adds another layer of meaning to linguistic content.

Additionally, it has been shown that music can convey meaning. Harmony has been found to act as a prime in local and global contexts. Timbre too has been shown to impact the processing of other musical elements, such as pitch and intonation, and to transmit meaning, in particular meaning about instrument identity.

The goal of surveying such research is to provide evidence to support the hypothesis that emotion conveyed through vocal timbre impacts emotional perception of meaning and emotion in sung words. By showing how vocal timbre may be considered similar to emotional tone of voice, and by extrapolating the findings of research into timbre perception to vocal timbre, this chapter has established that there is sufficient evidence in the existing literature to support this hypothesis.

In summary, it is evident that we draw on sounds and nonlinguistic cues every day in the creation and extraction of meaning. If we are so attuned to these cues in life, then we should also be attuned to them in music and, thus, also in popular vocal songs. Given the prominent role that vocal timbre plays in vocal songs, it seems likely that vocal timbre would also be contributing to the meaning we extract from sounds. In particular, emotional meaning.

4 Vocal Timbre Analysis: Resources, Challenges, and Overview

4.1 Introduction

Literature surveyed in Chapter 2 has shown that timbre has always been considered an important musical parameter, and that it has been used by music makers throughout history to convey meaning. Chapter 3 presented evidence to suggest that vocal timbre has the potential to be a highly evocative stimulus, possibly contributing to perception of emotional meaning in song. Given the importance and prevalence of timbre discussed in previous chapters, it is perhaps surprising that there is relatively little musicological analysis exploring this musical feature.

The present chapter examines the state of analytical techniques for timbre and vocal timbre. Throughout this chapter, it will become apparent that, although developments in vocal timbre analysis continue to be made, analytical techniques are yet to explore vocal timbre by considering its impact on emotional perception of lyrics.

Why focus the analysis of vocal timbre on popular vocal songs?

While vocal timbre has always been an important and inherent aspect of music, this thesis proposes analysing vocal timbre in popular vocal songs specifically. Before continuing, it is worthwhile articulating the two main reasons for this decision.

First, notation is often not the primary tool for composing popular vocal songs. Rather, this music tends to be documented and preserved through recordings. This opened up the potential for musical features to be used in new ways because they could now be recorded aurally, rather than only graphically. Vocal timbre is one such musical feature which benefited from this – the often "untrained" and characteristic qualities of individual singers' voices, which is often important to the sound of such songs, could now be preserved.

To elaborate further, historically vocal timbre has been temporal and circumstantial, existing only in the live performance. There did not exist for vocal timbre an accurate method of documentation (as there did for other musical elements such as melody). Recently, this has begun to change as technological developments now allow for the documentation of vocal timbre through magnetic and digital recording. Once recorded, a vocal timbre can also be heard repeatedly, with no variation. This allows an analyser/listener to become familiar with vocal nuances, and allows analysers to develop systematic methods of analysis (see section 4.3 for some examples).

Popular vocal song is one genre which has relied on these developments for their creation and dissemination. It is precisely the considerable reliance of popular vocal songs on these technological developments which makes them particularly suitable to study vocal timbre – a musical feature that also depends on these same developments to be systematically and efficiently analysed.

A second reason popular vocal song is preferred for analysing vocal timbre is the effect developments in amplification technologies had on increasing the ability of singers to exaggerate and project their vocal nuances. The microphone is a tool which can project a variety of vocal sounds at a range of dynamics. Using this tool, singers can exaggerate expressive aspects of a vocal timbre without being overpowered by other musical features (see Chapter 2 for a more detailed discussion). For example, a singer may now project vocal utterances such as grunts and breathiness, even if they are being accompanied by a full rock band. These technological developments are important tools in the creation and performance of popular vocal songs. Indeed, it is because of the use of such technology that, "[i]n popular music broadly considered, timbre is one of the most active parameters of experimentation, and a primary means of differentiation among artists and styles" (Heidemann, 2016, p. 2). In this way, popular vocal songs may be considered a good prism through which to examine nuances of vocal timbre.

In summary, although vocalists have, undoubtedly, always utilised timbre as part of their vocal performance, historically this has been difficult to document and thus to analyse. Today, technological developments allow for the documentation, study, manipulation, and control of timbre in new ways. These developments are also heavily used by composers and performers in the creation and dissemination of popular vocal songs. These ties between technological developments being instrumental in the preservation of vocal timbre and also being important tools in the creation and dissemination of popular vocal songs, make popular vocal songs a good resource through which to analyse vocal timbre.

4.2 Challenges of Analysing Vocal Timbre Through Popular Vocal Songs

Given the potential for popular songs to be used as a resource for analysing timbre, it is interesting to note that relatively little work has been done in this area. Although in recent decades popular music has begun to be increasingly included in musicological contexts (as will be explored in section 4.4), historically speaking, popular music has struggled to find acceptance as a subject of “serious” study (see, for example, Middleton, 1993; Seeger, 1977; Tagg, 2000 for a general discussion). The response to popular music from musicology was “more often than not . . . marked by insult, incomprehension or silence” (Middleton, 1993, p. 177). This led to popular music studies being largely marginalised in musicological contexts in the early- to-mid-20th century. Consequently, the potential for this music to be used to analyse vocal timbre was not realised. Three broad reasons may account for this early marginalisation: concepts of high- and low-brow art, a lack of appropriate analytical techniques which allow for recording based analysis, and a dismissive attitude in the mid nineth centuary towards the use of technology in musical creation and disemination.

4.2.1 The concept of high-brow versus low-brow art forms.

The idea that popular music was not "serious" enough to warrant analysis was a prevalent attitude throughout 20th century musicology (Tagg, 2000). In the early 1900s, for example, there was a tendency in traditional musicological circles within established academic institutions to view jazz as primitive and superficial, therefore not worthy of "serious" study. It is certainly the case that the creators and listeners of popular music itself have contributed to this divide. Consider, for instance, rock'n'roll in the 1950s. This music was, in part, the result of the younger generation rejecting traditional (i.e., Western art music) musical conventions as a symbol of their broader rejection of traditional values and norms (as well as musical conventions). However, the inability to completely understand or make sense of popular music through traditional Western approaches to analysis[10] also contributed to it being viewed as unworthy of study. In other words, popular vocal song was not worthy of serious study because it did not fit a predetermined, prescriptive idea of how musical meaning *should be created*.

[10] Such approaches tended to prioritise those musical features which can be documented easily and (relatively) accurately through conventional scores and graphic notation. Therefore, they are not well equipped to deal with some musical features of popular vocal song, including vocal timbre.

Charles Seeger describes this problem as the “failure to distinguish between the prescriptive and descriptive” (1977, p. 168) analytical approaches. According to Seeger, prescriptive analytical methods involve the application of a predefined model for how music should be composed/performed. Such a method is based on a preconceived set of ideas about what is “good” and “meaningful”. On the other hand, descriptive analytical approaches focus on describing a piece in-and-of-itself. That is, a descriptive approach provides a specific report of a piece rather than a report of how that piece conforms to a predetermined musical model.

Taking a prescriptive approach to popular vocal song analysis has two major negative impacts. First, this approach implies that a popular vocal song must be justifiable in terms of Western art music norms to be worthy of “serious” study. Prescriptive approaches tend to use only analytical techniques which are best suited to the analysis of musical features that can be represented through conventional scores and traditional notation techniques. Consequently, they tend to reflect the aesthetics of Western art music (which is primarily documented through these forms of notation). Therefore, a prescriptive approach implies that, for a popular vocal song to be considered “serious” music, its analysis should be possible using existing analytical techniques and thus comparable to Western art music. Prioritising the similarities between popular vocal song and Western art music, and the use of exclusively traditional analytical techniques, marginalises not only popular vocal song (which does not always fit a prescriptive model) but also vocal timbre, as vocal timbre cannot be accounted for through traditional analytical techniques alone.

Second, taking such an approach to popular vocal song analysis implies that meaning in these songs can only be understood and created by audiences and music makers who are “musically educated” (i.e., those with a formalised understanding of harmony and analysis in the Western art music sense). It was often asserted that popular music audiences in general were incapable of comprehending any deeper musical meaning (see, e.g., Cole, 1970; Mann, 1963). Similarly, the belief that one needs a formalised, traditional musical understanding to create “good” music can also be seen in discussions of popular music’s creators (see, for example, the discussion on jazz in Lambert, 1934). This approach is problematic as it implies that, if it is only those musical features foregrounded in traditional analytical techniques that make music “good”, then other musical features such as vocal timbre are, by comparison, meaningless. However, vocal timbre plays a considerable role in the creation of popular vocal songs. Therefore, if one dismisses popular vocal song because it is created or consumed by people with no traditional, formalised musical training (i.e., it is not a high-brow art form), then one also misses the opportunity to analyse vocal timbre through this music.

4.2.2 Recording-based analytical techniques.

Analysis’ “central activity is comparison” as this is how one “determines the structural elements and discovers the functions of those elements” (Bent, 1987, p. 5). To conduct such an analysis one requires an appropriate set of tools. However, these tools are not always available for vocal timbre.

Typically, vocal timbre analysis has received comparatively little attention from traditional musicology compared to other musical features such as melody and harmony. One reason for this may be that, historically, composers tended to explore elements such as harmony and melody to a greater degree than vocal timbre, with the interpretation of timbre and vocal timbre largely being left up to the performer. This may have been because these elements were more easily transmitted through conventional scores and graphic notation—one of the few ways to preserve music before the development of recordings. Musicologists too may have also chosen to study these elements because they were more easily preserved and, therefore, easier to analyse (after all, one can only study what has been documented and preserved). Today, the ability to record music has allowed vocal timbre (and individual performances and realisations of musical works) to be preserved and heard over and over. This has brought vocal timbre into greater focus, making it easier to study.

Nonetheless, it remains the case that vocal timbre still "resists description" (Heidemann, 2016, p. 1). In other words, vocal timbre remains a musical feature that is difficult to discuss—it is always performed (and therefore able to be captured on recordings), but is rarely represented on the page. This lack of conventional graphic representation for vocal timbre is problematic not because vocal timbre *must* be graphically represented, but because much of musicological analysis relies on traditional approaches to studying music. Such approaches tend to prioritise the analysis of musical features which can be notated. Therefore, these techniques are not well equipped to deal with musical features such as vocal timbre.

One example of the application of traditional analytical techniques marginalising vocal timbre can be found in harmonic analyses of popular vocal songs. Wilfrid Mellers (1973), for example, conducted a harmonic analysis of The Beatles' music in his book *Twilight of the Gods*. The Beatles are generally recognised as a band who wrote quite harmonically complex music. Therefore, Mellers' harmonic analyses do yield interesting results. However, by applying only traditional analytical techniques and focusing on only harmony, the importance of other, more performance based, musical features such as timbre are dismissed. The same is true for many other analyses of this time. For example, William Mann's (1963) analysis of The Beatles, which also analysed this music's harmonic structure. By focusing on harmonic analysis, these analyses prioritise the written score over other, more performance based, musical elements.

In popular music, prioritising the score and the written representation of music over the performance/recording can be problematic as it diminishes the importance of other musical features which are not conventionally notated (such as vocal timbre) (Middleton, 2000, p. 6). Certainly, it is unrealistic to expect analyses such as those discussed in this section to broach the full breadth of musical features present within a popular vocal song. Nevertheless, the general trend to ignore performance-based musical features in favour of those represented through conventional scores and traditional notation techniques has resulted in incomplete analyses of popular vocal songs (see, e.g., discussion in Moore, 2010, pp. 257–258). While projects such as the Centre for the History and Analysis of Recorded Music are making inroads into analyses of recordings (The AHRC Research Centre, 2010), much still remains to be done in the field of performance/recording based analysis (Auslander, 2004, pp. 1–10).

4.2.3 Dismissive attitude towards the use of technology.

As touched on in section 4.3.2, technological developments have allowed vocal timbre to be preserved and, therefore, to be more easily analysed. However, while this technology has facilitated the preservation and analysis of vocal timbre, and the development of popular vocal songs, it has also been the cause of some pushback from early 20th century scholars. Certainly, there were some scholars who embraced this new medium for documenting music, for example Seashore's *Psychology of Music,* but in general the dependence on recording/playback devices, and amplification, was seen as a symptom of popular vocal songs' inferiority to other forms of music. This was problematic as technology played a large part in the creation and dissemination of such music. Therefore, dismissing popular vocal song based on its reliance on technology meant that the opportunity to analyse vocal timbre through this genre of music was also missed.

One prominent justification for this dismissal was the belief by some scholars that audiences were not attracted to popular vocal songs in any meaningful way, rather they were aroused simply by the loudness of the music—loudness which was made possible through the use of loudspeakers (technology). For instance, in Mann's 1963 article, he asserts that "it is the sheer loudness that appeals to Beatles admirers" (Mann, 1963, p. 6). Similarly, Peter Cole, in his 1970 article entitled "Lyrics in Pop", states his belief that audiences of earlier popular music were attracted to popular songs because of the danceable rhythms and "screaming, shouting rag-tops", not because these songs had any deeper lyrical meanings:

> Pop is music, not poetry. It is not the words which sell records; it is the feeling, the impact a record makes on the listener. A great number of the pop fans who buy records never pay close attention to the words; they just demand that they should be there. It is very rare for a purely instrumental record to sell. So, paradoxically there is a demand for lyrics, for the presence of a human voice on the record, even though the lyrics are not listened to or cannot be heard. (Cole, 1970, p. 20)

Another justification given was that, because popular music was constantly being played through loudspeakers at audiences, technology encouraged a *passive* (i.e., superficial) kind of listening. For example, one could often hear popular music on the bus, in the shopping centre, at the train station. In these cases, the loudspeaker allowed popular music to invade public life and meant that popular music was never deeply engaged with by an audience (e.g., who analyses the harmony of a song when it is being played through the train station speakers?). Lambert is one scholar who held this view. In his 1934 discussion on "mechanised music" Lambert states that:

> It is obvious that second-rate mechanical music is the most suitable fare for those to whom musical experience is no more than a mere aural

> tickling, just as the prostitute provides the most suitable outlet for those to whom sexual experience is no more than the periodic removal of a recurring itch. The loud speaker is the street walker of music. (Lambert, 1934, p. 239)

In short, Lambert asserts that loudspeakers cheapen music because they are invasive, project music everywhere, leave no room for silence and, therefore, promote passive listening. Consequently, music associated with such mechanisation is low-brow and not worthy of in-depth study. Excluding popular vocal songs from "serious" musicological discussion due to their use of technology also excludes the analysis of vocal timbre (since this technology is the primary mode through which vocal timbre may be preserved and documented).

Criticisms of the relationship between popular music and technology prevailed in other forms of academia too. For example, in his 1938 essay "On the Fetish-character in Music and the Regression of Listening", Adorno criticises popular music for being merely a product of mass culture and points to mechanisation as a symptom of this. Adorno sees this mechanisation as degrading the music in favour of the object that produces it (Adorno, 1985). An example of this would be a person placing greater importance on wanting to own a nice radio, rather than on wanting to hear the music coming from the radio. In this way, both Lambert and Adorno criticise mechanisation because it reduces the meaning in music by overexposing audiences and prioritising the materialistic object (i.e., the loudspeaker) over the supposed meaningful one (i.e., the music). This sentiment is problematic for popular vocal song as mechanisation is important for its creation (recording studios, amplification, etc.) and consumption (playback devices like record players, which do not require the listener to "read" or understand music in any traditional, formalised way). Dismissing the primary method through which this music is produced and shared means that musical features such as vocal timbre (which also relies on this method) are also dismissed.

4.3 Approaches to Vocal Timbre Analysis Today

While the discipline of musicology may have been reluctant to analyse popular music in the early 20th century, other disciplines such as sociology were much more proactive in studying this music. In particular, these fields were extremely active in investigating popular music's social and cultural impacts. Consequently, there exists much more research on popular music as a *social phenomenon* than on popular music as a musical text (D. K. Blake, 2012, p. 1). This "historical lag" (Seeger, 1977, p. 168) has led to a lack of well-developed and widely agreed upon analytical techniques for many musical features of popular music, including vocal timbre. Today this is beginning to change as popular music has been increasingly included in musicological contexts in recent decades. This section will outline some of the key approaches to popular music, popular vocal song and vocal timbre analysis during this time. It will show that there remains a gap in analytical techniques that could account for both emotionally valenced vocal timbres and lyrics, while also highlighting those approaches that are beneficial to the form of vocal timbre analysis proposed in this thesis.

It is important to note here that this review is restricted to English texts. This is not to negate the important work in the field of the voice in popular music from non-Anglophone scholars, such as Catherine Rudent. Rather the restriction is pragmatic – without English translations it is difficult for the researcher (i.e. myself) to ascertain the meaning of such non-Anglophone texts.

4.3.1 New analytical approaches.

In recent decades, popular music in general has been increasingly recognised in academic contexts as a subject worthy of attention (Moore, 2003, pp. 1–2). This can be seen from the inclusion of popular music courses at universities, to the founding of academic journals dedicated to the field, to the establishment of popular music publications more generally (e.g., *Rolling Stone Magazine*) focused on the discussion and analysis of popular music. As part of this recognition, analyses of popular music have begun to be conducted which use traditional analytical approaches in descriptive ways. For example, Cooke investigates the relationship between phrase structure and harmonic progressions in his article "The Lennon–McCartney Songs" (Cooke, 1982). Another example of a descriptive analysis can be found in Burns's use of a modified, reductive, form of Schenkerian analysis.[11] Burns's goal was to demonstrate voice leading and harmonic features, as well as highlight how harmonic manipulations may be particular to specific song narratives and artists (Burns, 2000). Burns has also extended her use of Schenkerian analysis to explore how the relationship between harmony, lyrics, and performance can be used to convey metaphors of strength, weakness, hierarchy, and closure (Burns, 1997).

Harmonic and narrative form (Neal, 2000, 2007), rhythm (Temperley, 1999), melodic-harmonic divorce (Temperley, 2007), and the use of modified Schenkerian analysis (Forte, 1995) have also been adapted to conduct popular music analysis. While all these studies demonstrate how popular music in general may be approached from a descriptive standpoint, the potential for popular vocal songs to be used as a resource for analysing vocal timbre has not yet been fully realised.

One reason for this is that, in these analyses, there remains an emphasis on the written score. However, as has been discussed in section 4.3, the use of only notation can be problematic as it assumes "that the full auditory parameter of music can be represented by a partial visual parameter" (Seeger, 1977, p. 168). Indeed, in the case of musical features which are not documented using traditional notation techniques, "prioritizing visual representation" can negatively impact on an analysis as it plays down the

[11] Schenkerian analysis is a process whereby harmonies are gradually reduced to reveal an underlying skeleton structure of tonic–dominant–tonic.

importance of the “auditory sensation” (D. K. Blake, 2012, p. 3). Therefore, to analyse vocal timbre through popular vocal song, one needs to take an aural approach.

The need for aurally based analytical techniques has been shown by some scholars through the application of transcriptions. In these cases, the value of the transcription lies in the process of transcribing, not necessarily in the resulting graphic representation of the song. This is because the act of transcribing allows the analyser to engage in an in-depth way with the recording of the song. Such engagement may allow the analyser to uncover musical complexities which may otherwise have been missed (see, e.g., Brackett, 1995; Jairazbhoy, 1977).

Take, for instance, Winkler’s 1997 analysis “Writing Ghost Notes: The Poetics and Politics of Transcription”. Here, Winkler uses transcription as a tool for exploring the difficulties of applying notation to a music that has a largely aural tradition (Winkler, 1997). His analysis, while not providing an exact transcription of the song, is successful in demonstrating the ways in which the process of transcribing can be beneficial:

> Rather than seeing it as a way of distancing oneself from the music, transcription should be seen as a deep and intimate involvement in musical processes. The goal should not be an objective representation but just the opposite: transcription must be recognized as an intensely subjective act. And this should be recognized not as a fundamental weakness, but as a fundamental strength. (Winkler, 1997, p. 200)

Another example is Walser's 1995 article, "Rhythm, Rhyme, and Rhetoric in the Music of Public Enemy". Here, Walser uses the process of transcribing the rap song "Fight the Power" to highlight the "poverty" of notation in representing the specific sound of this song (Walser, 1995, p. 203). By doing so, Walser illustrates the particular complexities found within the music and demonstrates how transcriptions "[are] particularly useful . . . because cohesiveness and complexity are precisely what have been denied to hip hop, and those are the qualities that notation is best at illuminating" (Walser, 1995, pp. 199–200). In this way, the process of transcription, then, can be used in analysis to shine a spotlight on the specific complexities of popular music (Winkler, 1997, p. 170), and to demonstrate the "poverty" of notation and analytical techniques for musical features such as vocal timbre.

One technological development that has changed the way transcriptions can be used is the spectrograph—an electronic transcription that generates, rather than conventional notation, a visual image of sounding harmonics. However, the problem with such technology is that electronic transcriptions are often too objective, tending to generate too much information to be treated as a functional transcription (Cogan, 1987; Cogan & Escot, 1976; Jairazbhoy, 1977). Furthermore, if one wishes to generate a spectrograph of only a single specific musical feature within a song, one must have that musical feature as an isolated track—something that is not always possible. Therefore, while the use of spectrographs can be useful in analysis, more benefit may be derived from the process of engaging with the (audio and visual representations of the) recording, rather than from the spectrographs alone. In other words, spectrographs are “a means to an end rather than an end unto itself” (Reed, 2005, p. 11).

In summary, the studies cited in this section provide examples of how existing analytical tools can be applied to popular music analysis. However, these tools typically require modification (e.g., shifting focus to the process, rather than the output, of transcriptions). While modifying these tools for the analysis of other musical features has been beneficial, their application to vocal timbre analysis is limited (as, modified or not, traditional analytical techniques still tend to rely on the written score). The process of transcription, however, holds promise as spectrographs can be used to visually confirm what is heard aurally. Although one must always remember “that the full auditory parameter” cannot always be “represented by a partial visual parameter” (Seeger, 1977, p. 168) and that it is the process, rather than the product, which is often of most value.

4.3.2 Word-Music Analysis

In examining the relationship between music and lyrics in popular songs, this thesis is situated within a long tradition of word painting. Word painting refers to the use of musical gesture to illustrate the (literal or figurative) meaning of a word or phrase (Carter, 2001). Common examples of word painting include melodic lines imitating directional text, such as “down” or “fall”, “onomatopoeia (for example, the imitation of the sounds of battle, birdsong or chattering…) … figurative or pictorial melodic or contrapuntal gestures … and scoring (a single voice for ‘all alone’; three for the Trinity)” (Carter, 2001, para. 4). The practice has a long history, with examples of word painting being found from antiquity to modern day music (see, for example, Harran, 1986, as well as Strykowski, 2016, for a discussion of word painting in 16th century madrigals and chansons). As metaphor, word painting can be used to convey, subvert, or strengthen a lyrical message (see, for example, Kroeger, 1988; Zbikowski, 2009). Due to this multi-modal nature, word painting may be useful for understanding the underlying cognitive structures involved in the processing of music and language (Zbikowski, 2009).

Specifically in relation to popular vocal songs, the potential for musical features such as timbre and processing effects to be employed as a form of word painting have begun to be explored. In this vein, Serge Lacasse has explored how "the manipulation of voice through recording techniques can contribute to the mediation of… expressive moments" (Lacasse, 2010c, p. 211). One example is the relationship between lyrics and processing effects, such as reverberation. Lacasse identifies voiceless consonants sounds as occurring often in the verses of Peter Gabriel's "Blood of Eden" (Lacasse, 2010c, p. 217). The reverberation amplifies the high frequencies of these voiceless consonants, causing them to resonate for a long period of time after the word has ended. Lacasse identifies this relationship as a metaphor for reflection:

> In fact, right from the outset, there is an obvious relation between the idea of 'reflection' in the first line and the long reverberation following the word, a metaphor that is extended during the whole verse ('caught sight', 'saw', etc.) (Lacasse, 2010c, p. 217).

Lacasse also goes on to draw comparisons between this use of reverberation, and the potential for reverberation to signify love and religion

(Lacasse, 2010c, p. 217). By examining the impact of technology on the timbral experience of the lyrics, and by making suggestions as to what these timrbal qualities may signify, Lacasse is able to infer the possible implications of this word-music relationship on listener perception.

Aspects of word-painting from each of the above texts inform the analytical technique presented here. By including frequency and loudness in analysis (see section 7.3.1.2), this thesis recognises that more than one musical feature may be required to successfully imitate, or subvert, textual meaning through music (this is in line with Strykowski's 2016 finding). It is also suggested that, similar to Zbikowski (2009), shared cognitive structures may account for the success of multi-modal metaphors – in particular, how vocal timbre manipulations can be used as a metaphor for messages conveyed through lyrics.

By synthesising aspects of each of the approaches discussed above, and by suggesting a new analytical approach, this thesis can enrich the word-music tradition. In particular, word-music analyses conducted using the analytical technique developed in this thesis may offer new interpretations not only of popular music, but also of vocal timbre more generally. Further, as this analytical technique is focused on the relationship between the spoken and sung voice, and as much of popular music singing draws on this relationship (see, for example, Lacasse, 2010b, Middleton, 2000), this analytical technique is well positioned to be used in the analysis of large numbers of diverse popular songs. Finally, as this analytical technique is focused on a singer's vocal timbre (rather than a processing technique or manipulation), it may be applied to performance of many musical styles. For example, modern day recordings of madrigals and chansons, both of which are known for using word painting in relation to other musical features, such as melody.

4.3.3 Phonomusicology and other recording-based analytical techniques.

Studies which attempt to analyse popular music exclusively through the recording have, in the past, been quite few. One reason for this is that musicology has tended not to engage in performance analysis, and the field of performance analysis rarely engages with music, specifically popular music (Auslander, 2004, p. 1). Today, as technology which can assist with recording-based analysis develops (e.g., recordings are easily accessible, it is possible to splice/slow down recordings, we have access to spectrographs), the study of musical recordings is becoming increasingly common.

While many disciplines have benefited from the increasing availability of recordings, there are a few which would simply not exist without them. Popular music study is a field which relies heavily on recordings as a primary text of study (Cottrell, 2010; Katz, 2010). Such recording-centred study is commonly referred to as phonomusicology, a term which generally means "the study of recorded music, including its contexts of production and patterns of consumption" (Cottrell, 2010, p. 16).

One example of this kind of recording-based analysis can be found in David Brackett's *Interpreting Popular Music* (1995). Here, Brackett works primarily with the recording, using other tools to supplement the analysis. For example, in his analysis of Bing Crosby and Billie Holiday's "I'll Be Seeing You", Brackett uses transcriptions in the form of conventional notation and spectrographs to support his observations of what is heard aurally (Brackett, 1995, Chapter 2). Brackett's analysis covers a number of musical features, including vocal timbre. In this discussion of timbre, spectrographs are used to visually confirm what is heard aurally. However, the spectrographs used in Brackett's analysis make it difficult to distinguish specific vocal nuances. While this may be a criticism of the spectrographs used in Brackett's 1995 analysis, it is not necessarily one that holds true today. Today the analyser not only has access to computer software that can produce more detailed and clearer spectrographs, but they also have easier access to a variety of recordings and isolated vocal tracks. Should one today use spectrographs to support aural observations in the same way Brackett does, then one would be better equipped to produce clear and precise images. Such developments may also better equip one to discuss timbre in more detail, specifically in relation to its emotional meaning and relationship to lyrics (as proposed in this thesis).

Another instance of a methodology being developed solely to deal with the recorded text can be found in Allan F. Moore's *Song Means: Analysing and Interpreting Recorded Popular Song*. Here, Moore sets out a complete methodology for popular song analysis with the goal of uncovering "*how* they [songs] mean, and *the means* by which they mean" (Moore, 2012, p. 1). Two aspects of Moore's *Song Means* are of particular interest to this thesis.

First, the (inevitable?) use of interdisciplinary approaches in the formulation of a complete methodology for the analysis of popular songs: embodied cognition, the mirror neuron system, and psychology all contribute to Moore's methodology (see Moore, 2012, pp. 1–16). This lends credence to the interdisciplinary approach used in this thesis as it shows the increasing prevalence of nonmusical research in musical studies in general. Brackett's method, written in 1995, already hinted at the role of psychology and embodied cognition (i.e., the idea that we may draw on our physical experience to understand abstract concepts) in music research. Moore's method, written in 2012, has no reservations about overtly using such scientific approaches to the study of music. In this way, *Song Means* is a prime example of how the study of recorded music is not limited to the musicological sphere.

The second aspect of Moore's approach that is interesting for this thesis is that, as part of this methodology, Moore addresses both vocal timbre and lyrics. Timbre is recognised as an aspect of *shape*, shape being the *sound-world* of a recording which is made up of instrumentation and sound sources in general (see Moore, 2012, Chapter 2). Shape is important as it is "the sound world set up by a track that frequently forms the point of entry for a listener, that first triggers a sense of recognition" (Moore, 2012, p. 19). In other words, timbre is one of the most immediately recognisable musical features within a song. The voice, and consequently vocal timbre, may be particularly recognisable since, as Moore puts it, the distinction between popular song and popular music is the "interaction of everyday words and music" (Moore, 2012, p. 3). These "everyday words" rely on the voice to sing them, making the voice a musical feature which permeates our aural experience of popular songs. *Song Means*, however, does not offer a detailed discussion of vocal timbre. Rather, it is discussed as an aspect of timbre in general, or in relation to delivery (Moore, pp. 101–108). Lyrics, while being addressed, are also not discussed in detail in relation to their emotional meanings or their emotional relationship with vocal timbre. However, *Song Means* does highlight the need for an approach that can fuse the analysis of lyrics with musical analysis. This is because the analysis of popular songs tends, in Moore's words, to "address one [i.e., lyrics] at the expense of the other [i.e., musical features]" (Moore, 2012, p. 3). The present thesis bridges this gap by addressing both lyrics and vocal timbre simultaneously.

The development of analytical techniques in electroacoustic music is another example of how one may approach the study of recorded sound. While the subject matter is different (electroacoustic music vs. popular music), there are several processes used in this field which can inform vocal timbre analysis. This is primarily because electroacoustic music deals "with music that is not note-based and often lacking a representation equivalent" (Blackburn, 2009, p. 1). The same is true for vocal timbre, which is not preserved through conventional scores or typically notated in signs and symbols. Thus, electroacoustic music, like vocal timbre, must be analysed from the recording.

Several academics have made important inroads in recording based analysis. For example, Moylan (2007), Zagorski-Thomas (2014), Katz (2010), and Zak III (2001). However, one way this has been achieved in electroacoustic music analysis which is of particular interest to the methodology of this thesis is through spectromorphology[12] (see Schaeffer's *Traite des objets musicaux*", 1966). "The two parts of the term refer to the interaction between sound spectra (*spectro*-) and the ways they change and are shaped through time (-*morphology*)" (Smalley, 1997, p. 107). Spectromorphology is "concerned with perceiving and even thinking in terms of spectral energies and shapes in space, their behavior, their motion and growth processes, and their relative functions in a musical context" (Smalley, 1997, pp. 124–125). In other words, spectromorphology explores how sound can be understood in reference to its role in the music, its role in the human experience, and its relationship to other sounds.

A key aspect of spectromorphology is the use of "diagrammatic vocabulary sets". These sets are a tool to address "qualities that sounds inherently possess, comprised from commonly used language" (Blackburn, 2009, p. 1). They are useful for vocal timbre analysis as they retain emphasis on the aural sensation while also providing a way of describing sounds as they happen and evolve in time,[13] and in a way that is accessible to musicians and nonmusicians alike (something that is important in this and other popular music, as this music is not created only by classically trained musicians, and is often discussed and analysed outside of musicology). Blackburn expands on Smalley's concepts, proposing they may be used in the compositional and pedagogical teaching of electroacoustic music (Blackburn, 2009, p. 1). The

[12] Spectromorphology, a term coined by Denis Smalley based on theories developed by Pierre Shaeffer's.

[13] That sounds have motion is central to Smalley's spectromorphology and its application to electroacoustic music.

extension of spectromorphology in this way speaks to its practicality—if it can be used to teach purely aural music, then it must be equally well suited to analysing it. In this way, electroacoustic music analysis may be useful in analysing vocal timbre in popular music as it primarily deals with the analysis of the recording.

4.3.4 Semiotics.

The field of semiotics may be particularly useful for vocal timbre analysis for two reasons:

- Semiotics is concerned with how elements (including sounds) relate in the creation of meaning. This perspective is useful for the kind of vocal timbre analysis proposed in this thesis as it allows one to study vocal timbre by considering how vocal nuances relate to a listener's emotional experience (for some examples of a semiotic approach to music analysis see Lacasse, 2000; Nattiez, 1990; Tagg, 1999, 2000, 2011, 2012; Tagg & Clarida, 2003; van Leeuwen, 1999).
- A semiotic approach allows for the musical performance (or recording) to be the primary text of study.

Today, there are a number of well-known instances of semiotics being applied to music analysis. One such example can be found in Nattiez's 1990 book *Musical Discourse: Towards a Semiology of Music*. Here, Nattiez explores how:

> music-or any other social or cultural configuration-must be understood not only as a self-contained object, "a whole composed of 'structures' [but also] the procedures that have engendered it (acts of composition), and the procedures to which it gives rise: acts of interpretation and perception". (Nattiez, 1990, p. ix, as cited in Wood Massi, 1992, p. 1287)

Nattiez defines these acts as the immanent, the poietic, and the esthesic levels of music and applies them in his book to “investigate the question of ‘writing about music’” (Nattiez, 1990, p. 37).

A semiotic approach to music is also explored in Tagg and Clarida’s study *Ten Little Title Tunes: Towards a Musicology of the Mass Media*. The authors based their analyses of popular music on the general hypothesis that “commonly shared musical meanings (relations of musical signifiers to signified) do exist within a given culture and that they are formed and altered under particular social, ideological, technological and musical cultural contingencies” (Tagg & Clarida, 2003, p. 106).

Van Leeuwen, in his book *Speech, Music, Sound*, also explores meaning in music, focusing on the communicative nature of sounds with particular reference to popular music:

> In contemporary popular music sound matters more. Key singers and instrumentalists develop their own, immediately recognisable styles of singing and playing, and, thanks to recording, these can now become part of the language of music and be imitated and transformed by countless others. Saxophones can be soft and mellow, or tense and

> strident, like a hoarse whisper or a foghorn in the mist. Voices can be soft, smooth and well oiled, or rough, raspy and cracked. And singers as well as instrumentalists use a large repertoire of howls, wails, groans and other vocalisations. (van Leeuwen, 1999, p. 127)

As this quote illustrates, van Leeuwen places great emphasis on sounds and the ways in which they relate to audiences' real life experiences. Sounds are no longer simply by-products of actions, but rather they tell us something about the action and its role (van Leeuwen, 1999, p. 128). Another way this idea may be illustrated is through the use of *font* in graffiti. The font in which words are written has been used by artists as an expressive tool in graffiti such that, rather than being a means to an end, font is now contributing to the message. Similarly, the use of amplification when singing means that singers no longer have to project over other instruments (i.e., they no longer have to sing loudly) simply to be heard. Consequently, vocal utterances can now be used in ways that were previously impractical. Listeners may also now draw meaning from this sound by relating it to the human experience (e.g., the sound of the spoken voice, sounds heard in nature; see Chapter 3 for a more detailed discussion). In other words, sounds that may have previously not been audible/preservable, or been thought of as by-products, a means to an end, may now be used as expressive elements within themselves. To approach vocal timbre analysis through examining how vocal utterances may contribute to emotional meaning may be particularly fruitful as it "allows us to [fuse] ideological meaning and emotion" (van Leeuwen, 2012, p. 325).

Smith Alexander Reed takes a similar approach in his 2005 dissertation. Reed explores how vocal timbre may be examined through its links to our everyday, lived, experiences, and through our own experiences of making vocal utterances (Reed, 2005) (the notion that how we extrapolate meaning from these vocal utterances in music may be influenced by our own experience of making such sounds is explored in more detail in section 4.4.4.). This idea that sounds may be meaningful because of their connection to the human experience is particularly useful to draw upon in this thesis as it suggests that musical sounds do not always exist as "purely musical", but rather they may represent that which they embody and which they intend (D. K. Blake, 2012, p. 11).

The semiotic approach, while providing an avenue through which one may analyse popular music and vocal timbre, also allows the analyser to draw on further interdisciplinary techniques to support their analysis. One way in which semiotic analyses have been extended is through the use of reception tests to further explore the relationship between musical sounds and meaning. Reception tests are used in much of Tagg's work, for example. They can also be found in Lacasse's 2000 PhD dissertation which examined the role of vocal staging (roughly, the alteration of the voice through the production process with the intention of impacting listener perception; see Lacasse, 2000, p. 4, for a more detailed definition) on audience perception of recorded rock music (Lacasse, 2000). Lacasse's reception test showed that vocal staging impacts perception. The understanding about the role of vocal staging obtained in this way then informed his analysis of rock music. The need for these reception tests is essential to the success of analyses such as Lacasse's as they provide empirical evidence for what may otherwise be seen as “subjective” observations (Lacasse & Lefrancois, 2008, p. 3). The use of reception tests is an avenue that holds particular promise for the kind of vocal timbre analysis proposed in this thesis as the observations made in musical analyses presented in later chapters could otherwise be labelled as “subjective” observations.

In summary, a semiotic approach to vocal timbre analysis may be useful as it allows for analyses that:

1. are conducted with particular consideration of how musical elements are related in the creation of meaning;
2. can be supported by reception tests which are used to better understand the role these elements play in audience perception; and
3. place emphasis on the performance, using notation to assist with the analysis, rather than as the basis for analysis.

Further exploring the relationship between vocal timbre and meaning, as proposed in this thesis, may yield meaningful results not only for the field of analysis, but for better understanding vocal timbre perception in general.

4.3.5 Phenomenology and embodiment.

Several of the studies reviewed in section 4.4.2 and 4.4.3 (for example, Reed, 2005) suggest that sounds heard in music may be meaningful because of our own experience of producing such (if not exactly the same, then similar) sounds. If it is the case that a listener understands a sound by replicating within themselves the physical, emotional, or psychological state necessary to produce such sounds, then this may be an area of particular interest for vocal timbre analysis as, arguably, “popular singing acquires most of its inspiration from everyday speech” (Lacasse, 2010a, p. 226), and everyday speech is experienced almost universally. Therefore, listeners may have a well-developed frame of reference for interpreting, either consciously or nonconsciously, vocal utterances in popular vocal songs. Approaching music studies by considering how listeners may replicate musical sounds within themselves (physically, emotionally, or psychologically) generally takes two independent, but not necessarily mutually exclusive, forms: the phenomenological (i.e., more theoretical) line of enquiry, and the scientific (i.e., more empirical/quantitative) line of enquiry.

Phenomenology, while lacking a hard-and-fast definition, is generally concerned with “giving a direct description of our experience as it is” (Merleau-Ponty, 1945/1962, p. vii). It occurs when, “not content to ‘live’ or ‘relive’, we interrupt lived experience in order to signify it” (Ricoeur, 1981, p. 116, as cited in Godøy, 2011, p. 237). In this way, phenomenology may offer avenues for understanding the evocative nature of the sung voice in popular songs. While not dealing exclusively with vocal timbre, Clarke and Clarke’s *Music and Consciousness: Philosophical, Psychological, and Cultural Perspectives* (2011) is one example of how the concept of phenomenology can inform musical study. In particular, *Music and Consciousness* “begins the process of using critique to invent effective phenomenological methods for investigating the relationship between music and consciousness” (Gritten, 2013, p. 521). Another example can be found in Simon Frith's account of why the voice may be so expressive. Frith believes this expressive nature may be due to the sung voice being understood through the listener's "lived experience":

> The voice is a direct expression of the body, that is to say, is as important for the way we listen as for the way we interpret what we hear: we can sing along, reconstruct in fancy our own versions of songs, in ways we can't even fantasize instrumental technique – however hard we may try with our air guitars – because with singing, we feel we know what to do. We have bodies too, throats and stomachs and lungs. And even if we can't get the breathing right, the pitch, the note durations . . . we still feel we understand what the singer is doing in physical principle (this is another reason why the voice seems so directly expressive an instrument: it doesn't take thought to know how that vocal noise was made). (Frith, 1998, p. 192)

Similarly in his essay, "The Phonographic Voice: Paralinguistic Features and Phonographic Staging in Popular Music Singing", Serge Lacasse (2010, pp. 225–251) analyses the sung voice by considering listeners' lived experiences of the spoken voice. Lacasse explores how one may interpret vocal timbre in the sung voice in relation to para-linguistic features in spoken language, and uses this approach to describe the experience of vocal timbre in several case studies. Lacasse's study offers important findings relevant to this thesis as it explores how vocal nuances can impact perception of a song through their relationship with the spoken voice (i.e., their relationship with the everyday lived experience). Although it does not look specifically at the potential for these nuances to impact emotional perception of lyrics, it does lay the groundwork suggesting that such nuances are potentially meaningful and that they do contribute to musical expression and therefore potentially to lyric perception.

In “A System for Describing Vocal Timbre in Popular Song”, Kate Heidemann explores ways to give “a direct description of our experience as it is” (Merleau-Ponty, 1945/1962, p. vii) by presenting “a perception based system for describing vocal timbre” (2016, p. 1). In other words, Heidemann describes vocal timbre by articulating her own physical response to specific vocal sounds. Central to Heidemann’s proposed system is the concept of the “mimetic hypothesis” (Cox, 2011). The hypothesis is concerned with “how music becomes internalized into the bodies and minds of listeners” (Cox, 2011, p. 1). It is centred on the idea that one way through which to understand music is through examining how one imagines making the sounds they hear. An important aspect of this hypothesis is that this musical imagery may happen nonconsciously, that is, the listener may not be aware that this is how they are understanding the music. Such an approach allows the analyser to study a vocal timbre by describing the ways one may (potentially nonconsciously) imitate that vocal timbre through *mimetic motor imagery* (Cox, 2011, p. 2). Although Heidemann’s analyses and Cox’s hypothesis do not specifically explore emotion perception (as I do in this thesis), these studies, and those cited earlier in this section, do provide a good tool for studying how musical meaning can be understood—that is, through the bodily experience. My research takes this a step further, by exploring what that musical meaning may be.

4.4 Conclusion

This chapter has surveyed a broad range of literature and offered insights into the current state of analytical techniques for vocal timbre. On the whole, the treatment of popular vocal songs in musicology in the early part of the 20th century hindered vocal timbre analysis. The latter part of the 20th century, however, has seen vocal timbre being increasingly included in musicological study. Technological developments, specifically recording and playback devices, are a major factor in facilitating this inclusion as they allowed for vocal timbre to be preserved and heard over and over again. This has enabled more recording-based analyses to take place, which has, in turn, contributed to vocal timbre being increasingly studied.

A common thread in such studies is the use of interdisciplinary approaches. Very few of the studies cited in section 4.4 could be classified as belonging to only a single area of research. I believe that this interdisciplinary approach is the way forward for vocal timbre analysis—especially for the kind of vocal timbre analysis proposed in this thesis, as this analytical technique calls for:

1. The recording to be used as the basis of analysis such that emphasis may be retained on the aural sensation (drawing on phonomusicology);

2. Vocal timbre to be analysed in terms of its emotional impact on lyrics (i.e., looking at how musical features relate in the creation of emotional meaning—drawing on semiotics);
3. Emotion in vocal timbre to be described and classified through better understanding how listeners may internalise a heard vocal timbre within themselves (drawing on embodiment and phenomenology); and
4. The underlying hypothesis of this analytical technique, combining points 2 and 3, to be scientifically tested such that the musical analysis may be considered more than the idiosyncratic, subjective, observations of a single analyser (drawing on reception testing).

There is certainly ample room for the development of an approach to vocal timbre analysis as proposed in this thesis. Such an analysis should expand both our analytical techniques for vocal timbre, and the methodological tools for recording-based analysis in general.

Part II: Reception Tests

There is an old and deeply held tradition that vision “objectifies,” and, contrarily but not so widely noted, there is also tradition which holds that sound “personifies”. (Ihde, 2007, p. 21)

The emotional, physical and aesthetic value of a sound is linked not only to the causal explanation we attribute to it but also to its own qualities of timbre and texture, to its own personal vibration. So just as directors and cinematographers (even those who will never make abstract films) have everything to gain by refining their knowledge of visual materials and textures, we can similarly benefit from disciplined attention to the inherent qualities of sounds. (Chion, 2012, pp. 51–52)

5 Reception Testing: Investigating the Impact of Vocal Timbre on Sung Words

5.1 Introduction

In Part I of this thesis, it was established that vocal timbre is an important musical parameter that is likely to impact emotional perception of sung words. However, before proceeding to develop the proposed analytical technique on the basis of this premise, it is important to test this hypothesis. To this end, this chapter reports the results from reception tests designed to investigate whether vocal timbre (a) impacts emotional perception of sung words, and (b) whether it does so in an intersubjective way.

First a pretest was carried out to identify vocal timbres that would be clearly perceived as happy, neutral, and sad. Such a test was necessary as (a) vocal timbre can be considered highly subjective, and (b) no criteria or system of categorisation exist that allow emotional valence in vocal timbre to be easily and reliably identified. Vocal timbres identified in the pre-test were used in the subsequent test. This second test, the main test, draws on the happy and sad vocal timbres identified in the pretest to examine the hypothesis that emotion conveyed through vocal timbre impacts a listener's perception of emotion in sung words.

5.2 Pretest: Identifying the Emotional Valence of Vocal Timbres

5.2.1 Introduction.

This test was designed to identify vocal timbres that would be clearly perceived as happy, sad, or neutral for use in the main test. This was done by collecting listeners' ratings of a number of timbres which were supposedly happy, sad, or neutral and selecting the timbres considered to be the most characteristic of these emotional valences.

5.2.2 Methods.

Participants.

One hundred and nineteen participants (male = 53, female = 65, unknown = 1; 98% aged 18–29, 1% aged 30–39, 0% aged 40–49, 1% aged 50–59, 0% aged 60–69, 0% aged 70) were recruited from the University of New England and colleges using flyers and approaching participants in colleges and at Open Day events. Participation was voluntary. Participants who had a hearing impairment, who were not native speakers of English, and who had not spent most of their lives in Australia were excluded from the study (*N* = 1).

Materials and procedures.

The task was delivered as a survey where participants were required to listen to a musical stimuli sung in an emotionally valenced way, then rate those examples in terms of their emotional valence. Three emotionally valenced vocal timbres were tested: happy, neutral, and sad. For each timbral valence, six timbral tokens were selected. That is, six different kinds of happy vocal timbres (i.e., six different happy timbral tokens), six different kinds of neutral vocal timbres (i.e., six different neutral timbral tokens), and six different kinds of sad vocal timbres (i.e., six different sad timbral tokens). These happy, sad, and neutral timbral tokens were based on songs from the musical theatre genre (e.g. "I'll Cover You" from *Rent* (Dudescolded, 2010), "Windy City" from *Calamity Jane* (Friendofpoodles, 2008) and "Somewhere" from *West Side Story (*Ragna, 2012)), in which the vocal timbre of the singer was considered happy, neutral, or sad by virtue of its association with scenes whose role within the story was happy, neutral, or sad (e.g., a song in a mourning scene would be considered to have a sad timbre). Their purpose was to provide the singers with something to imitate.

The happy, sad, and neutral timbral tokens were recorded with three musical phrases in order to account for the possibility of melodic contours impacting participants' responses. These musical phrases were sung in crotchets, at a tempo of approximately 60 beats per minute. The three musical phrases were:

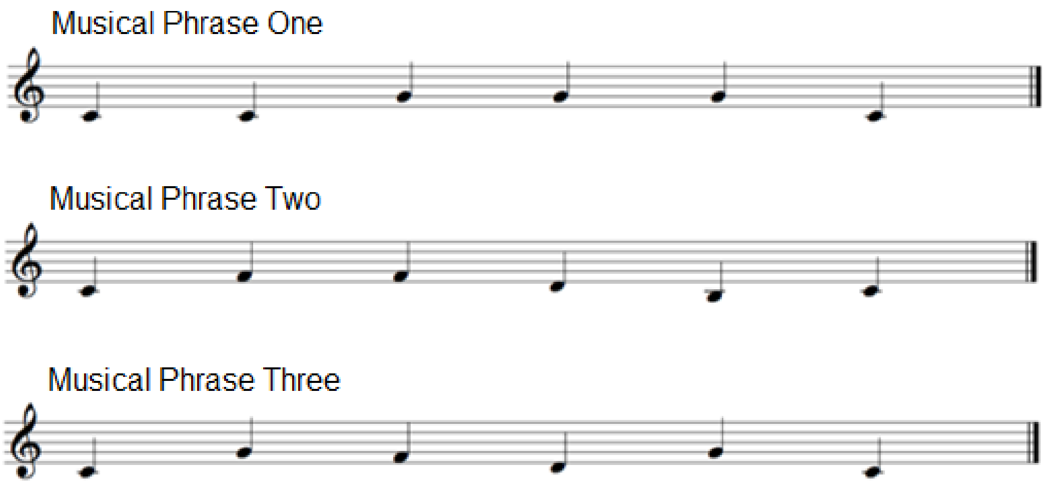


Figure 5.1. The three musical phrases used in the pretest.

To control for voice types and gender impacting judgments, three female and three male singers each recorded this set of musical stimuli, resulting in a total of 324 musical stimuli being recorded (18 timbral tokens (6 happy, 6 sad, and 6 neutral) x 3 melodic phrases x 6 singers (3 male, 3 female)).

Singers volunteered to help with the recording process. Unavoidably, small variations in speed and volume occurred across singers in some musical stimuli. Each musical stimulus had a duration of approximately seven seconds.

The 324 musical stimuli were counterbalanced across six lists on the basis of singer. This meant that each combination of timbre token and musical phrase would only appear once in each list, and each list would have that particular combination sung by a different singer, but all singers would appear an equal number of times in all lists. Therefore, each list contained all 54 token-phrase combinations (3 musical phrases sung with each of the 6 timbral tokens in each of the 3 valences) equally divided into male and female voices and with all singers being equally represented. Participants were randomly assigned to one of these lists. The musical stimuli were delivered using the Qualtrics (Qualtrics, 2005) survey platform through AudioSonic headphones. Participants were instructed to rate the emotional valence of each musical stimulus while working as quickly as possible. All lists began by providing examples of each valence; that is, musical examples of a happy, a neutral, and a sad timbre. These examples were not presented as experimental items within the list.

Scale.

Each participant rated the musical stimuli on a 5-point Likert scale. The scale was presented graphically starting with a large smiley face representing very happy, and ending with a large frowny face representing very sad. A score of 1 = very happy, 2 = happy, 3 =neutral, 4 = sad, 5 = very sad was assigned to each face.

5.2.3 Results.

Of the 6,396 total ratings, 608 were very happy (9.5%), 1,512 were happy (23.7%), 1,745 were neutral (27.4%), 1,568 were sad (24.6%), and 936 were very sad (14.7%). See Figures 5.2, 5.3, and 5.4 for an overview of ratings according to the attributes of interest: timbral token (Figure 5.2), musical phrase (the three different melodic phrases used) (Figure 5.3), and singer gender (male and female voices) (Figure 5.4).

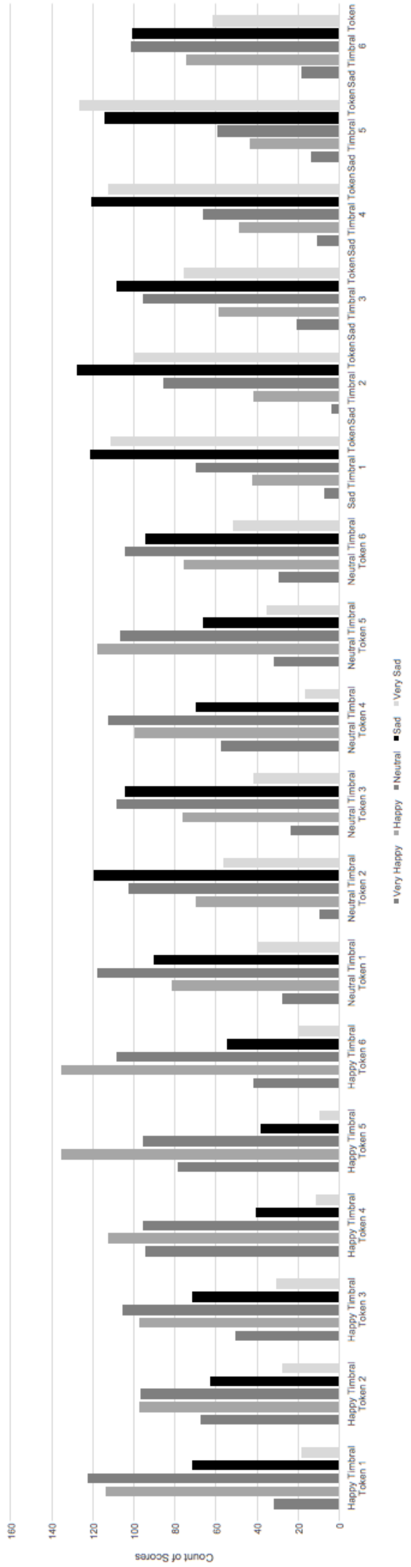


Figure 5.2. Distribution of stimulus ratings as very happy, happy, neutral, sad or very sad for each timbral token in each of the three emotional valences tested: happy, neutral, and sad.

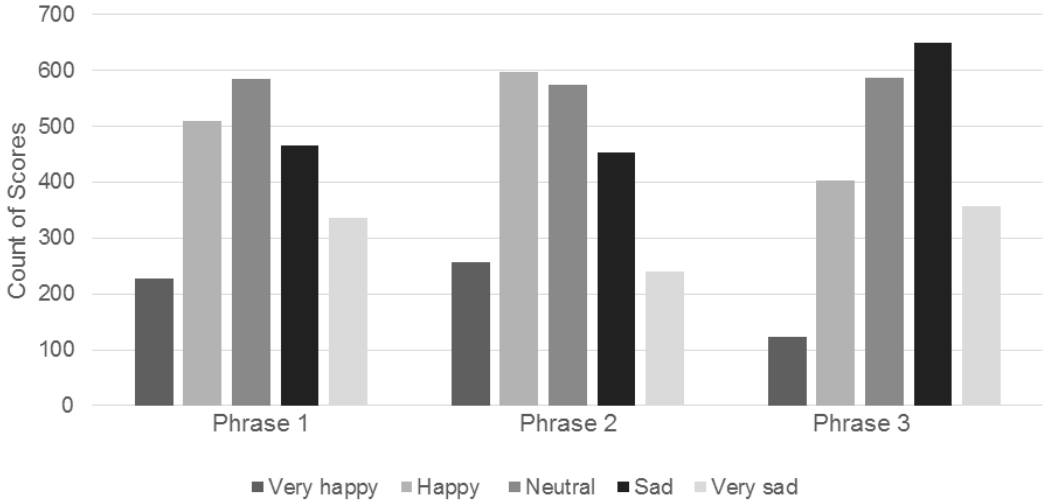


Figure 5.3. Distribution of stimulus ratings as very happy, happy, neutral, sad or very sad for each of the three musical phrases.

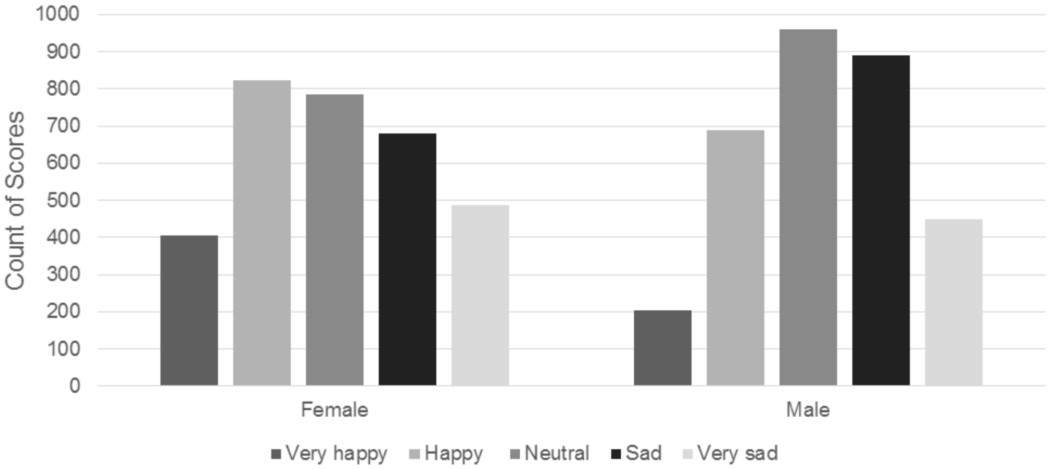


Figure 5.4. Distribution of stimulus ratings as very happy, happy, neutral, sad or very sad for the two singer genders.

To find out whether each of these three variables had any effect on participants' emotional valence ratings, three separate statistical tests were carried out. These tests compared participant's ratings along the emotional valence scale with respect to the three different attributes of the musical stimuli.

A one-way repeated measures analysis of variance (ANOVA) revealed significant differences between timbral tokens, F(5, 590) = 14.21 p < .001, η^2 = .11. This was expected as the timbral tokens were spread across three emotional valences: happy, sad, and neutral.

Additional pairwise comparisons were conducted to determine whether there were any differences between timbral tokens of the same emotional valence (i.e., pairwise comparisons between the happy timbral tokens, and between the sad timbral tokens; see Table 5.1). These tests revealed significant differences between some timbral tokens, however not all were significantly different from each other. For happy timbral tokens, there were no statistical differences between timbral token H4 and H5 (where the letter stands for emotional valence (i.e., happy), and the number stands for the particular token). These tokens were also shown to be considered "most happy" (H4 – M = 2.32, SD = .63; H5 – M = 2.34, SD = .65) (see Figure 5.5). For sad, pairwise comparison revealed that timbral tokens S1, S2, S4 and S5 were not significantly different from each other. Again, visual inspection showed these tokens to be considered "most sad" (S1 – M = 3.84, SD = .62; S2 – M = 3.79, SD = .67; S4 – M = 3.78, SD = .64; S5 – M = 3.85, SD = .72) (see Figure 5.5).

Table 5.1
Pairwise Comparisons Between Timbral Tokens (Happy, Sad, and Neutral)

Note. Pairwise comparisons between timbral tokens where the letter stands for emotional valence (i.e., H for happy), and the number stands for the particular token. The asterisk (*) indicates which pairs of timbral tokens are statistically different (e.g., H1 is different to H4, and H5).

Happy timbral tokens			Sad timbral tokens			Neutral timbral tokens		
H1	H2	*p* = .076	S1	S2	*p* = .555	N1	N2 *	*p* < .001
	H3	*p* = .985		S3 *	*p* < .001		N3	*p* = .323
	H4 *	*p* < .001		S4	*p* = .496		N4 *	*p* < .001
	H5 *	*p* < .001		S5	*p* = .855		N5 *	*p* = .002
	H6	*p* = .010		S6 *	*p* < .001		N6	*p* = .316
H2	H3	*p* = .060	S2	S3 *	*p* < .001	N2	N3 *	*p* = .006
	H4 *	*p* < .001		S4	*p* = .939		N4 *	*p* < .001
	H5 *	*p* < .001		S5	*p* = .465		N5 *	*p* < .001
	H6	*p* = .685		S6 *	*p* < .001		N6 *	*p* = .001
H3	H4 *	*p* < .001	S3	S4 *	*p* < .001	N3	N4 *	*p* < .001
	H5 *	*p* < .001		S5 *	*p* < .001		N5 *	*p* < .001
	H6	*p* = .056		S6	*p* = .163		N6	*p* = .910
H4	H5	*p* = .773	S4	S5	*p* = .327	N4	N5 *	*p* = .012
	H6 *	*p* < .001		S6 *	*p* < .001		N6 *	*p* < .001
H5	H6 *	*p* < .001	S5	S6 *	*p* < .001	N5	N6 *	*p* < .001

Although neutral timbral tokens did not need to be selected for use as experimental items in the main test, they will still be reported on here. For neutral tokens, pairwise comparisons showed timbral tokens N1, N3 and N6 to be similar, with N2, N4, and N5 each being statistically different from all others (see Table 5.1). Visual inspection showed that timbral token N1, N3, N6, and N5 were considered "most neutral" (N1 – M = 3.10, SD = .71; N3 – M = 3.19, SD = .70; N6 – M = 3.18, SD = .52, N5 – M = 2.88, SD = .57) (see Figure 5.5).

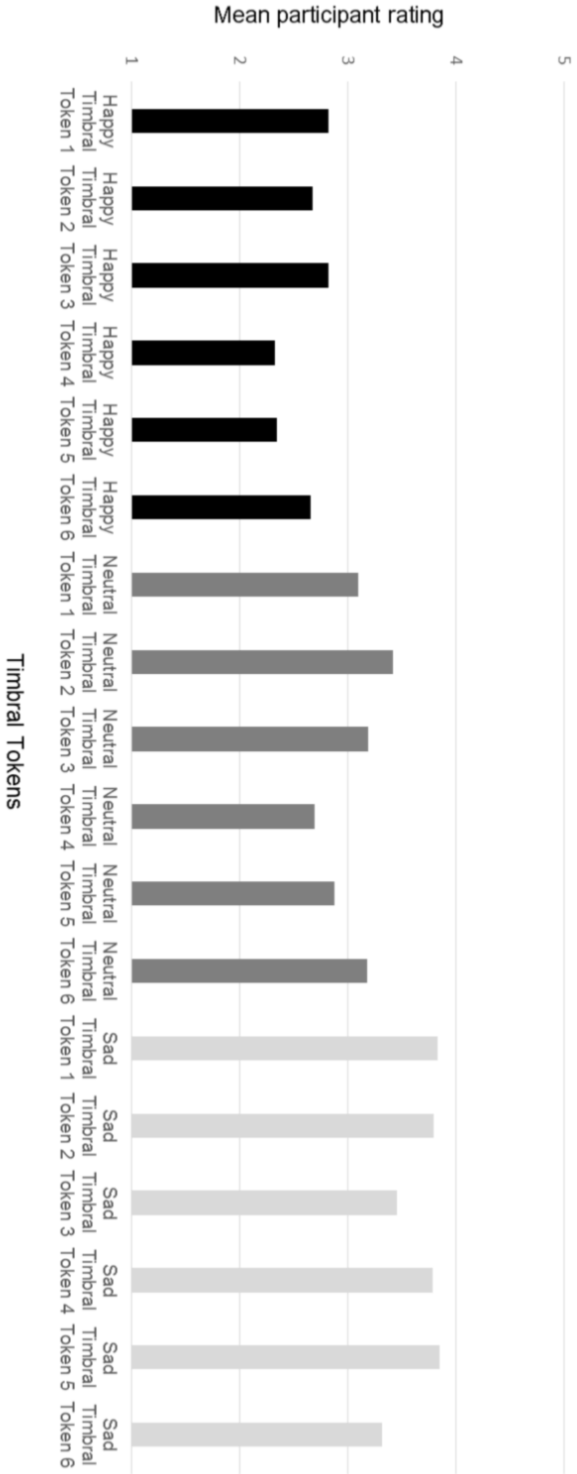


Figure 5.5. The mean participant rating for each emotional timbral token (1 = very happy, 2 = happy, 3 =neutral, 4 = sad, 5 = very sad).

A one-way repeated measures ANOVA revealed significant differences between musical phrases, F(2, 238) = 59.34, $p < .001$, η^2 = .33. Pairwise comparisons for musical phrases further revealed that all phrases were significantly different from each other within the three emotional valences (see Table 5.2).

Table 5.2

Pairwise Comparisons Between Musical Phrases Across the Three Emotional valences (Happy, Sad, and Neutral)

Note. Pairwise comparisons between musical phrases across the three emotional valences. The asterisk (*) indicates which pairs of musical phrases are statistically different.

Happy valences			Sad valences			Neutral valences		
Musical Phrase 1	Musical Phrase 2 *	$p < .001$	Musical Phrase 1	Musical Phrase 2 *	$p < .001$	Musical Phrase 1	Musical Phrase 2 *	$p < .001$
	Musical Phrase 3 *	$p < .001$		Musical Phrase 3 *	$p < .001$		Musical Phrase 3 *	$p < .001$
Musical Phrase 2	Musical Phrase 3 *	$p < .001$	Musical Phrase 2	Musical Phrase 3 *	$p < .001$	Musical Phrase 2	Musical Phrase 3 *	$p < .001$

Visual inspection revealed that Musical Phrase 2 tended to be considered “most happy” (M = 2.45, SD = .51), Musical Phrase 1 “most neutral” (M = 2.96, SD = .37), and Musical Phrase 3 “most sad” (M = 3.86, SD = .49). Visual inspection also revealed Musical Phrase 1 to have a slightly better spread between emotional valences than Musical Phrases 2 and 3 (see Figure 5.6).

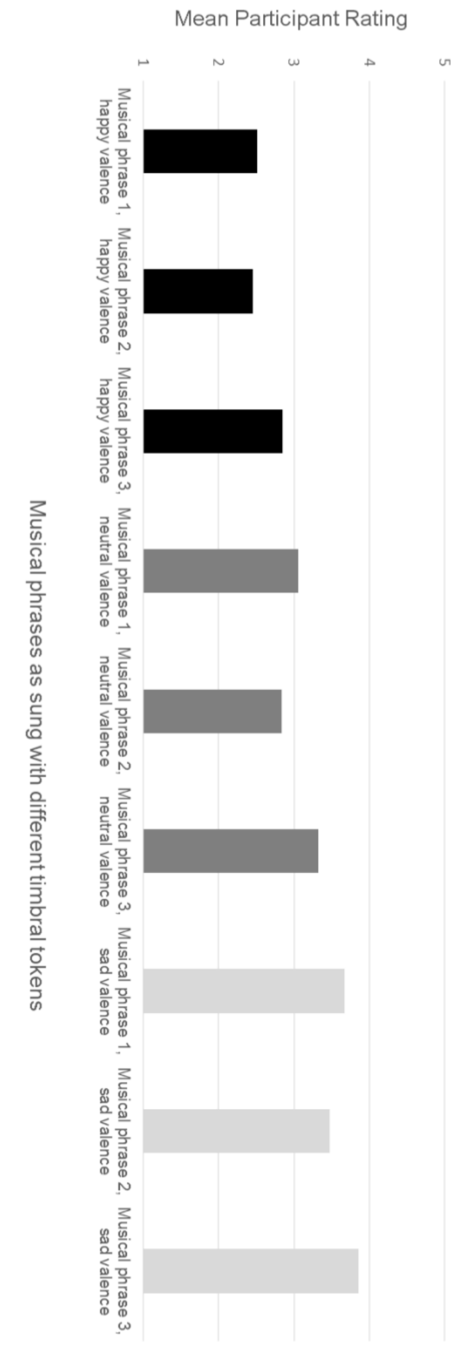


Figure 5.6. The mean participant rating for each musical phrase across the timbral valences of happy, neutral, and sad (1 = very happy, 2 = happy, 3 =neutral, 4 = sad, 5 = very sad).

A one-way repeated measures ANOVA also revealed significant differences between singer gender, F(1, 119) = 37.80, p < .001, η^2 = .24. T-tests were used to determine whether there were significant differences between singer gender across the three emotional valences (see Table 5.3). Significant differences were found for singer gender across the neutral and sad timbral tokens. Upon visual inspection, females were considered “most neutral” (M = 3.1, SD = .71), and males were considered “most sad” (M = 3.83, SD = .47). No significant differences were found for singer gender across the happy timbral tokens. Overall, visual inspection revealed that male singers demonstrated the widest spread between emotional valences (see Figure 5.7).

Table 5.3

T-test Results for Singer Gender Comparisons Across the Three Emotional valences (Happy, Sad, and Neutral)

Happy timbral tokens		**Sad timbral tokens**		**Neutral timbral tokens**	
Female	*Male*	*Female*	*Male*	*Female*	*Male*
M = 2.56 SD = .39	M = 2.65 SD = .53	M = 2.96 SD = .38	M = 3.19 SD = .45	M = 3.51 SD = .45	M = 3.83 SD = .47
t(236) = -1.40 *p = .16*		*t*(236) = -4.29 *p < .001*		*t*(236) = -5.36 *p < .001*	

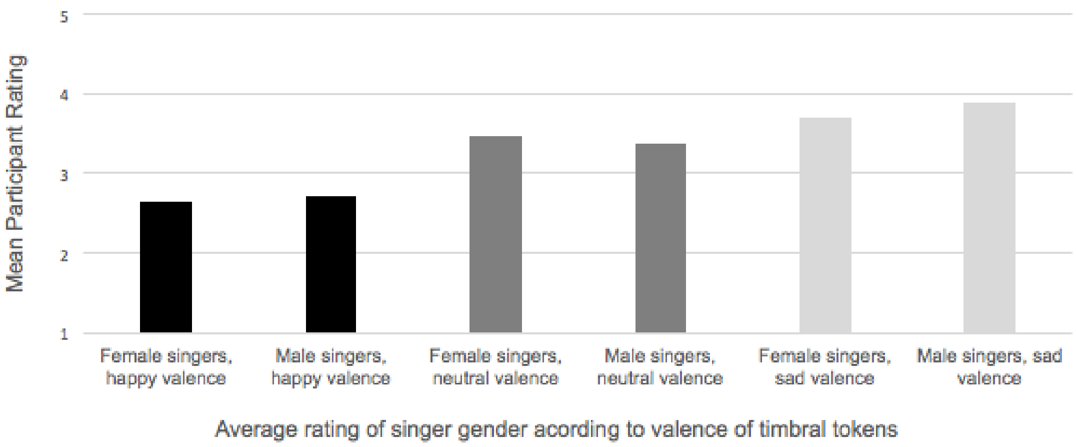


Figure 5.7. The mean participant rating for singer gender across the timbral emotional valences of happy, neutral, and sad (1 = very happy, 2 = happy, 3 =neutral, 4 = sad, 5 = very sad).

5.2.4 Discussion.

The purpose of the pretest was to identify emotional timbral tokens that were considered most typically happy and sad. The vocal timbre of these timbral tokens would then be used in developing the stimuli for the main test. It was found that, for each emotional valence, some timbral tokens were considered better exemplars of that emotional valence (e.g., some happy timbral tokens were considered significantly happier than others). In particular, happy timbral tokens H4 and H5 were not significantly different from each other and were also shown to be rated most happy (see Figure 5.5). Therefore, both timbral tokens would be suitable for use in the main test, and timbral token H5 was selected. In the sad emotional valence, timbral tokens S1, S2, S4 and S5 were not significantly different from each other and they were shown to be rated most sad. Given no statistical difference was found between these timbral tokens, any would have been suitable for use in the main test. Here, token S5 was selected. Neutral timbral tokens were also included in the pretest to ensure a fair range of emotional valences. However, although they will be included in the main test too, they will act only as filler items. Therefore, timbral token N5 was selected for use in developing the main test stimuli since it was one of the neutral tokens shown to be rated most characteristically neutral.

Similarly to the timbral tokens, significant differences were found between the three phrases across emotional valences. However, as would have been desired, no one phrase was shown to be considered both most happy and most sad depending on the timbral valence with which it was sung (Phrase 2 = "most happy" (M = 2.45, SD = .50), Phrase 1 = "most neutral" (M = 2.95, SD = .37), Phrase 3 = "most sad" (M = 3.85, SD = 0.49)). In principle, thus, all three phrases are equally suitable for use in the main test. What's more, phrases in the pretest had to be kept short (6 beats long) due to the large number of musical stimuli to be presented, but phrases in the main test needed to be longer to accommodate the phrase to be sung (10–13 beats long). Therefore, given that no one phrase was considered to be both most happy and most sad, and that the phrase for the main test could be longer than those used in the pretest, the main test phrase was designed to be a mixture of all three pretest phrases.

In order to keep the stimuli set at a manageable size in the main test, it was thought preferable to use only one singer to record stimuli. Statistical differences were found between genders for the neutral and sad valences, but not for the happy valences. Given that male singers were rated sadder than female singers, and that there was no difference between male and female singers in the happy valences, a singer of the male gender was selected for use in the main test (see Figure 5.6).

It should be noted that choosing the stimuli for the main test to be sung by a male singer is not to negate the potential impact singer gender has on listener perception. However, investigating the impact of singer gender on listener perception is not the aim of this research. Rather, this thesis is concerned with the impact of a singer's vocal timbre on perception—regardless of gender. The male singer gender was selected here as it was shown (by its slightly larger spread) to be the better tool for investigating the impact of vocal timbre on perception across emotional valences. While testing for differences in perception based on singer gender is beyond the scope of this research, this is another line of enquiry worth investigating in future research.

5.3 Main Test

5.3.1 Introduction.

The purpose of the main test was to test the hypothesis that underlies the proposed analytical technique: that emotion conveyed through vocal timbre impacts emotional perception of sung words. To this end, participants were asked to judge the emotional meaning of words in the presence of different emotionally valenced vocal timbres (those identified in section 5.2.4 of the pretest discussed above). The timbres' emotional valences either coincided or conflicted with those of the words. It was reasoned that, if the emotional charge of vocal timbre has the capacity to affect how sung words are perceived, participants should be faster and more accurate at identifying the emotional valence of words when they are sung in an emotionally matched vocal timbre than when they are sung in an emotionally mismatched vocal timbre.

The main test is modelled after the priming paradigm commonly found in studies into emotional tone of voice. In such experiments, the participant is first presented with an emotional tone of voice (i.e., a "prime" that is meant to bias the participant), then presented with a congruent or incongruent emotional word to which they have to respond. This is the same process used in the current study, except participants are presented with an emotionally valenced, sung, vocal timbre rather than the spoken voice. In the case of spoken words, the priming paradigm has revealed that emotional tone of voice does have the potential to impact perception of spoken words (e.g., Nygaard & Lundersl, 2002; Nygaard & Quees, 2008) (for a more in-depth review of this literature, see Chapter 3).

5.3.2 Method.

Participants.

Twenty participants (male = 8, female = 12; 70% aged 18–30, 20% aged 31–40, 10% aged 41–50) from the general population (i.e., who were not required to be musically trained or have any specialised musical knowledge) were recruited from the University of New England and surrounding community. All participants were native speakers of English, had spent most of their lives in Australia, and did not have a hearing impairment. Participation was voluntary. Participants were paid AU$15.

Procedures and materials

The study was conducted as a priming task in which the reaction times and response accuracy to different target words in different conditions was measured. Participants were required to listen to a musical phrase sang with timbres of different emotional valences and containing an emotional (either happy or sad) word (see Appendix). Their task consisted of identifying the emotional valence of the words as fast as possible.

The words, 56 in total, were selected from the *Affective Norms for English Words (ANEW): Instruction Manual and Affective Ratings* (Bradley & Lang, 1999) (see Appendix). This data set consists of approximately 600 words rated on a 9-point scale on three dimensions: pleasure (happy vs. unhappy), arousal (excited vs. calm), and dominance (controlled vs. in-control). For the current study, 18 happy and 18 sad words were selected as experimental items based on their ratings in the pleasure dimension (as this dimension deals with emotional valence). The average emotional valence of the selected happy words was 8.2 out of 9 (with higher scores representing happier valences, unlike in the scales used in these reception tests), ranging from 7.74 to 8.72. The average emotional valence of the sad words was 2.07 out of 9, ranging from 1.61 to 2.13. It was not possible to match the words on all the lexical parameters known to affect lexical retrieval. However, since all the relevant contrasts were within-item (i.e., reaction times and accuracy rates for happy words sung with a happy vocal timbre were compared with reaction times and accuracy rates for those same happy words sung in a sad vocal timbre, and the same was true in the case of sad words), it was not critical to do so. Additionally, 20 words with a neutral valence were used as fillers (mean emotional valence = 4.54 out of 9; range = 4.15 to 4.83).

The specific vocal timbre tokens to be used were determined in the pretest described in section 5.2.4 above. The chosen tokens (one happy and one sad) were recorded by a male singer as male singer gender was shown to give the best spread of scores across the three emotional valences (out of 5, mean happy score = 2.65, mean sad score = 3.19).

The pitches in which the singer sang for all recordings can be seen in Figure 5.8. All happy and sad words were recorded, together with the phrase "the word I will now sing is ____", both with a happy and with a sad vocal timbre (i.e., happy word and happy vocal timbre, same happy word and sad vocal timbre, and so on) such that a total of 72 experimental items were recorded (36 words x 2 timbral valences). For these target words, the emotional valence of the vocal timbre would thus either accord or disaccord with the emotional valence of the word itself. Additionally, the set of 20 neutral words were only recorded with a neutral vocal timbre and used as fillers to provide a greater emotional valence range.

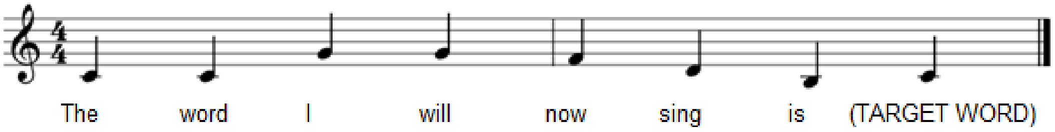


Figure 5.8. The musical phrase sung with the words "the word I will now sing is ____".

The stimuli so created were later tested against the selected stimuli from the pretest to check whether the timbres were the same as those they were supposed to replicate. This comparison test used the same perceptual paradigm used in the pre- and main tests. Eight experienced musicians were asked to rate how similar the happy and sad vocal timbres created for the main reception test (i.e., the timbres used as *experimental stimuli*) were to the chosen happy, neutral and sad examples from the pretest. This comparison test showed that the sad vocal timbres were perceived as being most similar to the sad example (74%), but the happy vocal timbres were less clearly identified with the happy example (31%), being often assimilated to the neutral one (53%). This shows an asymmetry in timbral emotional valence that may be inbuilt rather than an artefact of the way the materials were prepared: basically, for listeners, sadness seems to be more salient than happiness (in a similar way to negative emotions tending to be more sailent than positive ones; Rozin & Royzman, 2001). This asymmetry will be discussed later in more detail. Given the nature of the task (i.e., choosing between "happy" and "sad" only, with no possibility of choosing "neutral"), the fact that the happy vocal timbres were unlikely to be assimilated to the sad example (18% of the time), and that sad vocal timbres were unlikely to be assimilated to the happy example (4% of the time), was thought to be sufficient to test the differential effect of the timbre's emotion. However, the difference between the "happiness" of the happy timbre and the "sadness" of the sad timbre will be taken into account when interpreting the results.

Experimental items were counterbalanced across two lists on the basis of match between emotional valences of timbre and word (that is, a word appeared with a happy timbre in one list and with a sad timbre in the other) such that each list contained all the experimental words only once. There were equal numbers of each item type (i.e., happy word-happy timbre, happy word-sad timbre, etc.) in each list. Fillers were constant across lists. In this way, a total of 56 items were presented in each list. Half of the participants were randomly assigned to one list, and half to the other list.

The musical phrases were delivered in a different random order for each participant using the DMDX experimental program (Foster & Foster, 2015) through AudioSonic headphones. Stimuli were only presented aurally. Participants were instructed to indicate if the target word was happy or sad by selecting a key on the keyboard. The emotional valences were represented on the keyboard with a large smiley face representing happy on one key (*F* on the keyboard), and a large sad face representing sad on another key (*J* on the keyboard). To control for handedness, the order of happy and sad keys was reversed for half the participants in each list. Participants could only indicate happy or sad responses; therefore, it was expected that neutral items (fillers) should be randomly categorised as happy or sad. Participants were told to keep their fingers on the keys at all times. Reaction times were measured from the onset of the target word.

5.3.3 Results

Figures 5.9 and 5.10 show the reaction times and number of errors, respectively, of these within word comparisons according to the emotional valence of the word (hence forth referred to as word valence for short) and condition.

Two separate 2 (happy vs. sad words) x 2 (happy vs. sad timbres) repeated measures analyses of variance (ANOVA) were used to see whether there were differences in either reaction times or error rates for emotionally valenced words sung with a matched or a mismatched emotionally valenced vocal timbre.

In the case of reaction times (RTs), there was no main effect of timbral valence as responses to sad and happy timbres overall were not significantly different (F (1, 19) = 1.8, p = .195, η^2 = .09); and no main effect of word valence (F (1, 19) = .43, p = .59, η^2 = .02), as responses to sad and happy words overall were also not significantly different. Contrary to expectation, the interaction between word valence and emotional valence of timbre (hence forth referred to as timbre valence for short) was not significant (F (1, 19) = 2.49, p = .13, η^2 = .11). That is, words sung with an emotionally matched vocal timbre were not found to be responded to significantly faster than words sung with an emotionally mismatched vocal timbre. Visual inspection does, however, show a trend in the right direction, namely, a delay in word valence recognition in mismatched conditions. This delay appears more striking in the case of sad words (see Figure 5.9).

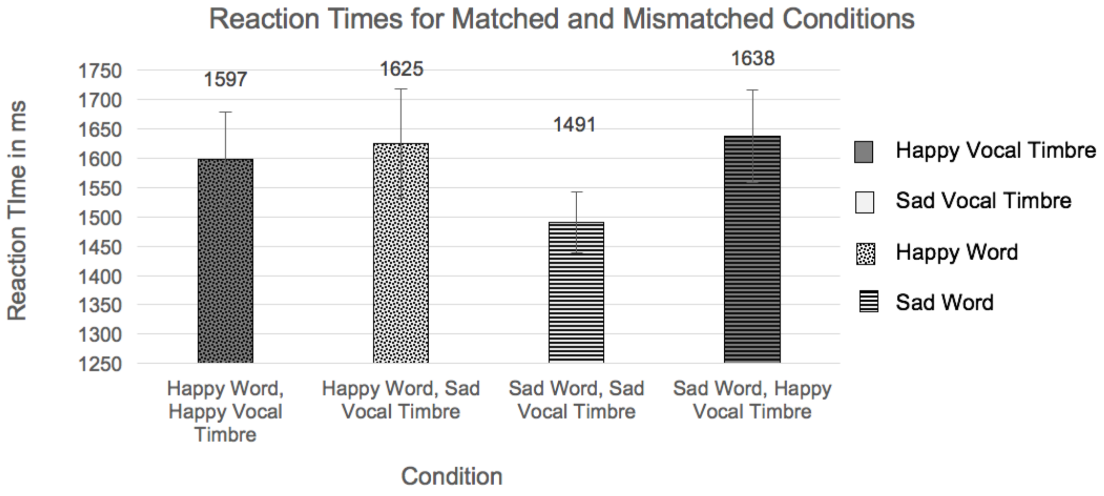


Figure 5.9. Reaction times for matched and mismatched conditions (and standard error bars). Emotional valence of the vocal timbres is indicated by shading of the background (white – sad, shaded – happy); emotional valence of the words is indicated by different patterns (stripes – sad, dots – happy).

Pairwise comparisons were used to further investigate response times to the two types of words separately. Results showed that response times to sad words when sung with a happy vocal timbre were significantly longer as compared to response times to sad words sung with a sad vocal timbre (t (19) = 2.48, p = .02). On the other hand, response times to happy words sung with a happy vocal timbre were not significantly different to those sung with a sad vocal timbre (t (19) = -.35, p = .73).

In the case of response accuracy (i.e., hitting the 'happy' key after a happy word such as paradise, and the 'sad' key after a sad word such as poison), there was no main effect of timbral valence (F (1, 19) = .37, p = .55, η^2 = .01), or of word valence (F (1, 19) = 1.69, p = .2, η^2 = .08). That is, participants were just as accurate *overall* in associating the word with the right emotional valence for both timbral valences, and just as accurate for sad as for happy words. Crucially, as expected, a significant interaction was found between word valence and timbre valence (F (1, 19) = 25.51, p < .001, η^2 = .54) indicating that significantly more errors were made for emotional words sung with a mismatched emotional timbre than for words sung with a matched emotional timbre (see Figure 5.10). That is, participants were more likely to associate a word with the wrong emotional valence when it had been sung with a vocal timbre of a different emotional valence.

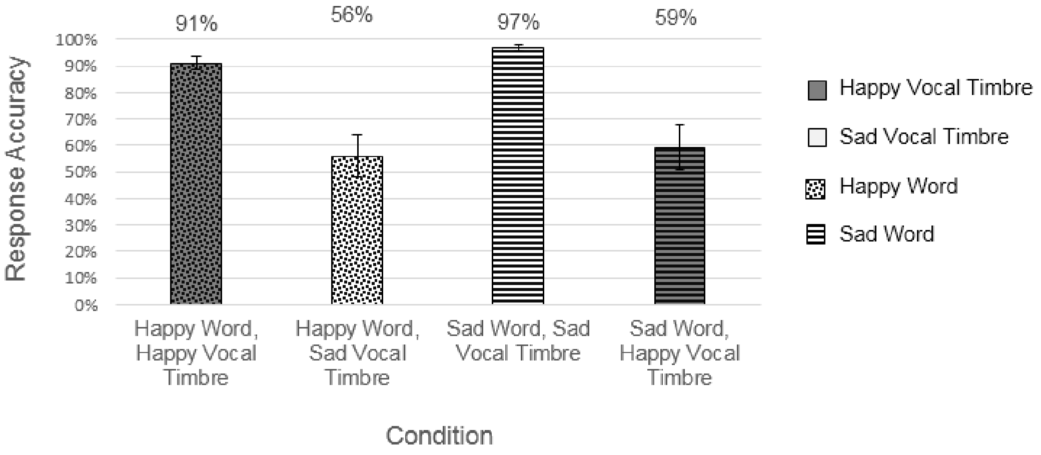


Figure 5.10. Percentage of correct responses for matched and mismatched conditions (and standard error bars). Emotional valence of the vocal timbres is indicated by shading of the background (white – sad, shaded – happy); emotional valence of the words is indicated by different patterns (stripes – sad, dots – happy).

Also as expected, neutral filler items were responded to randomly with 49% of neutral items being responded to as happy, and 51% being responded to as sad.

5.3.4 Discussion.

This reception test investigated whether vocal timbre could act as a prime, impacting on emotional perception of sung words. The results show that emotionally valenced words were recognised more accurately when sung with an emotionally matched vocal timbre than with an emotionally mismatched one. They also show that sad words were recognised faster when sung with a matching sad timbre. On the other hand, happy emotionally valenced words were not recognised faster when sung with an emotionally congruent vocal timbre. As a whole, these findings of an impact of nonlinguistic features (vocal timbre in this case) on perception of sung words conforms to expectations and are in-line with the results from the emotional tone of voice priming paradigm.

5.3.4.1 Emotional valence and word categorisation speed.

While these results do show the predicted effect of vocal timbre on the speed at which the emotional valence of sung words is recognised, the effect seems to be restricted to sad words. This result may not be as odd as it seems. Already in the comparison test performed by expert musicians, it was observed that happy vocal timbres were not considered as clearly happy as sad vocal timbres were sad: it was found that the sad vocal timbres used in the final reception test were most accurately identified as being similar to the sad vocal timbre they were replicating (74% of the time as opposed to 31% for happy words). In principle, it could be due either to the sad timbre offered as an example being easier to replicate for the singer or to the sad timbres so replicated being easier to identify for the musicians that performed the subsequent comparison task. In either case, given that these were all experienced musicians, it points to the possibility that sad vocal timbres contain more salient characteristics than happy vocal timbres (for example, more throaty, rough, breathy sounds). In fact, sadness and tenderness/love have been identified as one of the most expressive qualities in music performance (Juslin, 2001, p. 317). If so, it should not be so surprising that the effect on reaction times was most marked for sad words also in the general population.

In speech, negative prosodies have also been found to be more salient than positive ones in relation to accuracy in word discrimination: "independent of emotional meaning, words with happy prosody were rated less accurately than words with neutral or angry prosody" (Schirmer & Kotz, 2003, p. 1139); clearly, negative/sad emotional prosodies/vocal timbres are highly salient. Therefore, it is likely that these emotions are easier to categorise and may also impact word perception to a greater extent than their positive, high arousal counterparts (for example, happy prosodies/vocal timbres), as has been found here.

The reason why negative valences such as sadness appear to be more salient and more easily identified is likely to be related to the phenomenon known as negative bias. Negative bias has been observed for some time (see, for example, Hansen & Hansen, 1988; and discussion in the introduction to Rozin & Royzman, 2001, p. 296–298). It refers to the "asymmetry in the way in which people handle evaluatively positive and negative phenomena" (Lewicka, Czapinski, & Peeters, 1992, p. 425), with a tendency for negative experiences (thoughts, emotions, etc.) to be felt more acutely than positive ones (Baumeister, Bratslavsky, Finkenauer, & Vohs, 2001, p. 323). This negative bias has been explained as the result of evolutionary pressures which would benefit increased awareness to potentially dangerous stimuli (Hansen & Hansen, 1988; Rozin & Royzman, 2001).

In the case of this reception test, it may be that participants are impacted to a greater extent by the sad vocal timbre, intensifying its effect on sad words. This would explain why participants were quickest (by over 100 ms) in identifying word valence when both word and vocal timbre were sad (i.e., negative) (see Figure 5.9). At any rate, the salience of sad vocal timbres observed here is in line with research already cited in this section which suggests that negative emotional signals are likely to have more of an impact on emotional perception than positive ones.

An explanation for the lack of happy word RT effect as due to a less salient emotional valence of the happy timbre together with the pattern of results (see Figure 5.9) suggests the RT differences between the matched and mismatched emotional valence conditions could be the result of the congruent timbral valence facilitating recognition of the word's valence: the more recognisable sad timbre is thus associated with a larger effect on the congruent sad words. If the effect was the result of the incongruent timbres interfering with the recognition of the word's valence, it should have been the sad words that were less affected when being sung with a less incongruent happy timbre. However, in the absence of a neutral timbral condition, we cannot extract a definitive conclusion in this regard.

In sum, it is clear that timbral valence can influence the speed at which words are processed, at least for some words and in some circumstances. In particular, it is the more negative valences that show the greatest effect, and this may have to do both with sad timbres possibly being more salient and with a more general cognitive negative bias.

The mechanism by which such an effect takes place is beyond the purview of this study. However, we can speculate that sad timbres may predispose the listener to a state of mind in which the perceived likelihood and magnitude of a negative experience occurring is heightened (Rozin & Royzman, 2001). This negative emotionally valenced vocal timbre may heighten the listeners' sensitivity to further cues for "sadness", which may in turn facilitate the episodic trace of compatible sad words making their emotional valence easier to recognise (in a similar vein to Krestar & McLennan's (2013) findings).

5.3.4.2 Emotional valence and word categorisation accuracy.

As well as an effect on categorisation speed, an effect was also found for timbral valence on word categorisation accuracy: people were more likely to make mistakes when the timbral valence mismatched that of the word. Although errors in such an easy task are not necessarily expected, one explanation for this effect may be that participants traded response accuracy for speed. This is a well-known trade-off that occurs when people have to work under time pressure: participants can sacrifice accuracy to work as quickly as possible, or sacrifice speed to work more accurately (Rinkenauer, Osman, Ulrich, Müller-Gethmann, & Mattes, 2004; Wickelgren, 1977). As participants were instructed to work as quickly as possible in this experiment, this may have resulted in a sacrifice of accuracy for the benefit of speed.

But why would words be mis-categorised at all? One possibility is that participants were (sometimes) responding to the emotional valence detected in the timbre instead of that of the word. The emotion conveyed in the vocal timbre may create an expectation of the emotional valence of accompanying words that, when violated in mismatch cases, is more likely to elicit an incorrect response. That is to say, participants may be so influenced by the vocal timbre and the emotion conveyed therein that, in mismatched conditions, a conflict is induced between responding to the emotion elicited by the timbre and that evoked by the meaning of the word itself.

Alternatively, the effect on accuracy may be due to the emotional charge of the timbre imbuing the word itself and making it “feel” more (or less) happy or sad than it would in normal circumstances. This would, in turn, make the word less readily identified as happy or sad when sung with a mismatched emotional timbre and, thus, more easily misidentified. In other words, this effect could be due to the emotional valence of the words being altered to accord with the emotional valence created by the timbre.

Curiously, despite the asymmetry in timbral valences (sad was sadder than happy was happy) and in contrasts with the results of reaction times discussed in the previous section, the tendency for mistaken categorisation was the same for happy and sad words. One possibility is that what mattered in this case was whether the timbre was perceived as sad or not sad rather than sad or happy. The happy timbres used in the final reception task may not have been consistently found to resemble the happy timbre exemplar (31% of the time), but they were quite consistently recognised as not sad (82% of the time; similar to how consistently, 74% of the time, the sad timbres were recognised as clearly sad). It is possible that, for the purposes of reacting quickly, it was the congruent timbre that was responsible for most of the effect by putting the listener in the right emotional state to activate compatible episodic word traces while, for the purpose of categorising, perceiving the timbre as either clearly sad or clearly not sad may have been enough to bias word responses. In that case, a happy word with a sad timbre would be as difficult to categorise as a sad word with a non-sad timbre.

The explanations offered for the congruency effect (timbre valence creating expectation and timbre valence imbuing the word meaning itself) would imply quite a strong pull of the emotional valence of vocal timbre if it can compete with a word's intrinsic valence. What is clear, in any event, is that vocal timbre is contributing to the listener's perception of the word's emotional meaning and it does so in a consistent manner across this population of young- to middle-aged Australians. This is just what was predicted. Therefore, that participants were more likely to misidentify word valences in mismatched cases across both happy and sad words supports the hypothesis that a timbre's associated emotional valence affects the semantic processing of emotional words.

5.4 Conclusion

These results demonstrate that listeners do perceive emotional valence in vocal timbre, and this emotional valence can impact a listener's emotional perception of sung words in an intersubjective way. These results are significant for the analysis of vocal timbre in popular vocal song because they reveal the intimate relation between timbre and lyrics and point to the need for timbre analysis to take this relationship into account. Therefore, these results not only support the hypothesis that underlies my proposed analytical technique (i.e., that vocal timbre impacts emotional perception of sung words) but also highlight the need for such an approach to timbre.

In these reception tests, I have tried to control the conditions of the stimuli as much as possible in order to be able to investigate just those dimensions that were of interest, namely, the emotional valence of timbres and words. However, while this increases the validity and reliability of the results, it can also be thought to affect the ecological validity of the stimuli. Therefore, in the future, several further studies could be conducted to gain a more ecologically valid understanding of how vocal timbre impacts emotional perception of lyrics. A logical next step would be to test the role of vocal timbre in its wider musical environment (i.e., within the context of the entire song). Future studies should also consider the impact of repeated listenings on the emotional experience of a popular vocal song—i.e., how does our understanding of a song change as we grow more familiar with it? Also of interest is the impact of listening environment—i.e., how does the way we listen impact our experience of a song (and vocal timbre and lyrics)?

One further consideration is how one may efficiently and reliably identify emotional valence in vocal timbre in the context of a musical analysis. In this case, a pretest was run to identify happy and sad vocal timbres. However, this is not an efficient way to conduct a musical analysis. The objective analysis of the acoustic properties of emotionally valenced vocal timbres could be one way forward in this area. This is precisely what will be explored in the next chapter.

6 Potential Mechanisms for Identifying Vocal Timbre Valence

6.1 Introduction

The reception tests presented in the previous chapter have shown that listeners perceive emotional content in vocal timbre (with some types of emotional content being perceived better than others; i.e., sad vocal timbre seems to be more salient than happy vocal timbre), and that this emotional valence can impact a listener's emotional perception of sung words. These results suggest that it should be possible to analyse timbre on the basis of its emotional content. However, there still remains an important question to address: when conducting an analysis of vocal timbre in popular vocal songs, how can the analyser know whether a vocal timbre is happy or sad (or any other emotional valence)? It is important to find a robust and efficient way to identify emotional valence in vocal timbre since completing a reception test to identify the emotional valence of a vocal timbre for every musical analysis would be too time consuming and render the analytical technique redundant. This chapter presents a solution to such a problem by suggesting that one way to classify emotional valence in vocal timbre is through objectively analysing its acoustic features.

6.2 Developing the Timbral Attributes and Assessing Their Predictive Potential

An unplanned observation made when preparing for the pretest (see Chapter 5) suggested such an objective analysis of vocal timbre could indeed assist in the classification of emotional valence. The stimuli for the pretest consisted of vocal timbres that could be considered happy or sad (or neutral). These stimuli were recorded by several different singers who modelled their "happy" and "sad" vocal timbres off those found in musical theatre songs (musical theatre was used as the song's place and role in the plot could help to ascertain its emotional content). When recording the stimuli, there were no formalised preconceptions as to what would make these vocal timbres happy or sad. However, it was soon noticed that the form of specific acoustic features within a vocal timbre did seem to characterise certain emotionally valenced timbral. The acoustic features, which will henceforth be referred to as Timbral Attributes (TAs), that were identified in the pretest were: onset, sustain[14], termination, articulation, and contrast.

In the early stages of the pretest, it was recognised that these TAs could be useful in assisting singers emulating the timbres to be recorded. Therefore, the TAs were defined on a scale from their very happiest to very saddest forms (see Table 6.1) and used to aid in the stimuli recording process. For example, a

[14] When first developing these attributes, sustain was initially referred to as intensity. However, as "sustain" is an already well established term for referring to the way in which a note is held, it is now used in place of "intensity".

TA onset with an emotional valence of very happy is defined as "Mostly clear onsets. Beginning on the pitch". Conversely, a very sad onset here is defined as having "Less clear onsets. Notes tend to start airy or as noise and pitch is added after the initial onset".

Table 6.1

Scale Based on Timbral Attributes According to Their Happiness/Sadness

<table>
<tr><th>Timbral Attributes (TAs) at the very happy end of the scale</th><th>Timbral Attributes (TAs) at the very sad end of the scale</th></tr>
<tr><td colspan="2"><i>Very Happy</i> <i>Happy</i> <i>Neutral</i> <i>Sad</i> <i>Very Sad</i></td></tr>
<tr><td>Onset:</td><td>Onset:</td></tr>
<tr><td>Mostly clear onsets. Beginning on the pitch.</td><td>Less clear onsets. Notes tend to start airy or as noisy and pitch is added after initial onset.</td></tr>
<tr><td>Articulation:</td><td>Articulation:</td></tr>
<tr><td>Each note/word strongly re articulated, with separation between timbral sounds</td><td>Each note/word less clearly articulated and more continuity of timbral sound.</td></tr>
<tr><td>Sustain:</td><td>Sustain:</td></tr>
<tr><td>Full, strong timbre.</td><td>Less strong timbre, more airy and chesty sounds.</td></tr>
<tr><td>Contrast:</td><td>Contrast:</td></tr>
<tr><td>Belting—a big, full vocal sound.</td><td>“Building Timbre”—build anticipation/emotional interest by contrasting subdued with open timbres.</td></tr>
<tr><td>Termination:</td><td>Termination:</td></tr>
<tr><td>Singing through to end of note</td><td>Tendency to taper off at the end. As the termination tapers off, more breathy and unpitched sounds become present.</td></tr>
</table>

After the stimuli (i.e., the timbral tokens) were recorded, the analyser (i.e., myself) assessed each TA of each stimulus on the same 5-point scale that the participants would later use to rate the overall stimulus. TAs were rated individually such that one timbral token might have a TA rating of: onset–happy, termination–happy, contrast–neutral, sustain–very happy, articulation–happy. Using the same scale as the participants would later use allowed for easier comparison between identified TA emotional valence against participants' overall stimuli ratings. In other words, it allowed for the reliability of each TA as a predictor[15] to be easily assessed. Evaluating the predictive value of each TA was achieved by comparing the rating given to each of the stimuli's TAs by the analyser (i.e., myself) with participants' overall rating of the stimuli. For example, to determine if onset was a good predictor, the analyser would look at whether stimuli with sad onsets were more likely to be rated sad overall, and stimuli with happy onsets were more likely to be rated happy overall.

Figures 6.1–6.5 show the percentage of overall stimuli ratings in the pretest plotted against each TA's original rating. For example, Figure 6.1 shows participants' overall ratings of stimuli as sad, neutral, or happy, against the analyser rating of their onsets as sad, neutral, or happy. Before continuing

[15] A predictor is something defined independent of a listener's expressed perception that can reliably indicate (predict) what the listener will report perceiving.

on to discuss in more detail the reliability of each TA as a predictor, there are two points of note.

First, in the pretest, participants were able to rate each stimuli across a five point scale: very happy, happy, neutral, sad, and very sad. It was also across this 5-point scale that each TA was rated by the analyser. However, Figures 6.1–6.5 have been compressed from a 5-point scale to a 3-point scale (collapsing very happy and happy, and very sad and sad). This was done due to a tendency by participants to avoid selecting extremes when rating the emotional valence of timbre (henceforth referred to as timbre valence for short). A bias against selecting extremes of a scale, called central tendency bias, is one form of response bias that has been known for some time (Douven, 2018, p. 1203). Therefore, presenting a full 5-point scale may in fact hamper the visualisation of the trends of interest: whether TAs rated as happy/sad correlate with an overall happy/sad rating for that stimuli. In either case (5- or 3-point scale), a large number of neutral responses should be expected given the effect of response bias on participants' ratings.

Second, because the pretest was not designed to test whether these TAs would be good predictors, there was no counterbalancing of TAs to ensure all emotional valences and combinations of emotional valences were equally represented. Therefore, some TAs' emotional valences may have occurred more often than others. To improve comparability, Figures 6.1–6.5 are based on percentages. The total number of times each TA emotional valence occurred across the stimuli and in which combinations with other TAs' emotional valences can be seen in Table 6.2.

Table 6.2

Number of Stimuli in Which Different Combinations (in Pairs) of TA Emotional valences Could Be Found

Note. This table shows how many stimuli were rated as very happy, happy, neutral, sad, and very sad for each combination of TAs. For example, there were 36 stimuli with a rating of very happy for onset; of these, 13 stimuli were rated as having a sustain rating of very happy, 20 as having a sustain of happy, 2 as having neutral sustain, and 1 as having sad sustain.

	Onset Very Happy	Onset Happy	Onset Neutral	Onset Sad	Onset Very Sad	Intensity Very Happy	Intensity Happy	Intensity Neutral	Intensity Sad	Intensity Very Sad	Termination Very Happy	Termination Happy	Termination Neutral	Termination Sad	Termination Very Sad	Contrast Very Happy	Contrast Happy	Contrast Neutral	Contrast Sad	Contrast Very Sad	Articulation Very Happy	Articulation Happy	Articulation Neutral	Articulation Sad	Articulation Very Sad
Onset Very Happy	36																								
Onset Happy		44																							
Onset Neutral			136																						
Onset Sad				90																					
Onset Very Sad					18																				
Intensity Very Happy	13	0	0	0	0	13																			
Intensity Happy	20	27	27	2	0		76																		
Intensity Neutral	2	12	69	30	0			113																	
Intensity Sad	1	5	33	53	9				101																
Intensity Very Sad	0	0	7	5	9					21															
Termination Very Happy	30	6	3	0	0	13	24	2	0	0	39														
Termination Happy	6	33	21	2	0	0	36	9	14	3		62													
Termination Neutral	0	4	100	46	3	0	16	88	43	6			153												
Termination Sad	0	1	12	38	15	0	0	12	42	12				66											
Termination Very Sad	0	0	0	4	0	0	0	2	2	0					4										
Contrast Very Happy	13	0	0	0	0	10	3	0	0	0	13	0	0	0	0	13									
Contrast Happy	18	22	11	0	0	3	41	7	0	0	18	24	9	0	0		51								
Contrast Neutral	5	21	118	72	5	0	32	99	78	12	8	35	143	35	0			221							
Contrast Sad	0	1	7	13	13	0	0	2	23	9	0	3	1	26	4				34						
Contrast Very Sad	0	0	0	5	0	0	0	5	0	0	0	0	0	5	0					5					
Articulation Very Happy	28	3	5	0	0	13	16	2	5	0	25	7	1	3	0	13	11	12	0	0	36				
Articulation Happy	8	29	22	6	0	0	45	12	8	0	14	34	15	2	0	0	34	30	1	0		65			
Articulation Neutral	0	12	105	67	9	0	15	89	71	18	0	21	133	35	4	0	6	159	28	0			193		
Articulation Sad	0	0	4	14	6	0	0	7	17	0	0	0	4	20	0	0	0	17	2	5				24	
Articulation Very Sad	0	0	0	3	3	0	0	3	0	3	0	0	0	6	0	0	0	3	3	0					6

6.2.1 The predictive potential of each independent TA.

In the case of onset (Figure 5.1), it appears that stimuli that had been considered (i.e., rated by the analyser) to have a sad onset were quite consistently rated as sad overall by pretest participants. Similarly, across the neutral dimension, onset ratings of neutral correlate with neutral ratings of the stimuli overall. On the other hand, while stimuli considered to have a happy onset are also far more often identified as being happy than sad, these stimuli were just as likely to be rated neutral as to be rated happy by pretest participants. That is, whereas sad onsets have a very clear correlation with sad overall ratings of the stimuli, happy onsets tend to be considered neutral almost as much as they are considered happy. The response bias discussed in the previous paragraphs may partly account for this. However, another reason for this result may be the way onset is defined (see Table 6.1): sad onsets are very distinctive, whereas happy onsets may be more similar to neutral ones. This distinctiveness (perhaps itself another case of the negative bias described in the previous chapter) may allow for sad onsets to be more easily identified by the analyser in comparison with happy and neutral onsets, resulting in onset appearing to be a better predictor in sad cases. At any rate, the conclusion is that onset appears to be a good predictor overall, and an especially good predictor of sad valences.

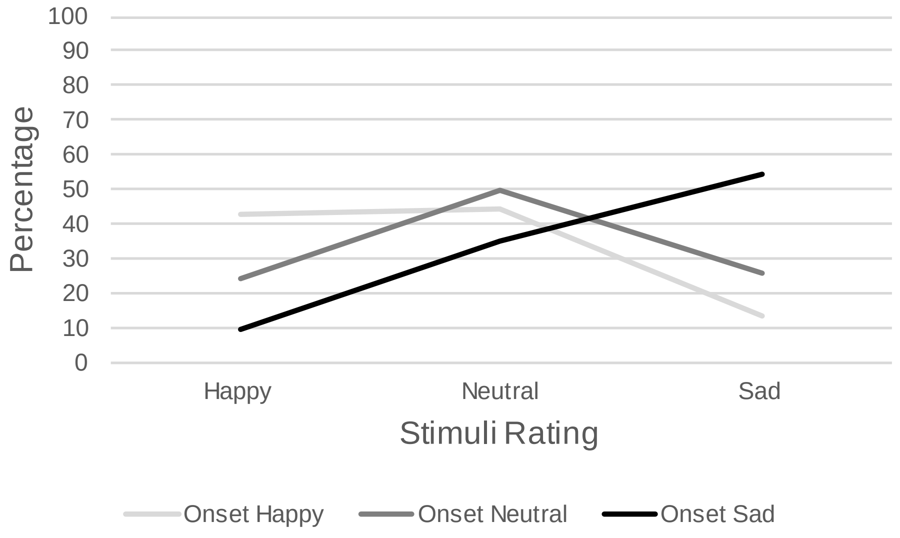


Figure 6.1. Percentage of happy (very happy and happy), neutral or sad (very sad and sad) ratings given to stimuli as a function of their onset emotional valence.

Sustain (Figure 6.2) is like onset in that sad, neutral, and happy TA ratings appear to correlate quite well with participants' ratings of the stimuli overall. Interestingly, both onset and Sustain present strong correlations in the neutral dimension despite the expectation, as discussed in section 6.2, of a high rate of neutral overall stimuli ratings due to known response biases towards the middle of scales (Furnham, 1986, p. 385). However, the trend evident in both the onset and Sustain TAs—happy, neutral, and sad TA ratings correlating with overall happy, neutral, and sad stimuli ratings—does not appear to be substantially affected by this bias. This suggests that these two TAs may act as reliable predictors in identifying emotional valence in vocal timbre.

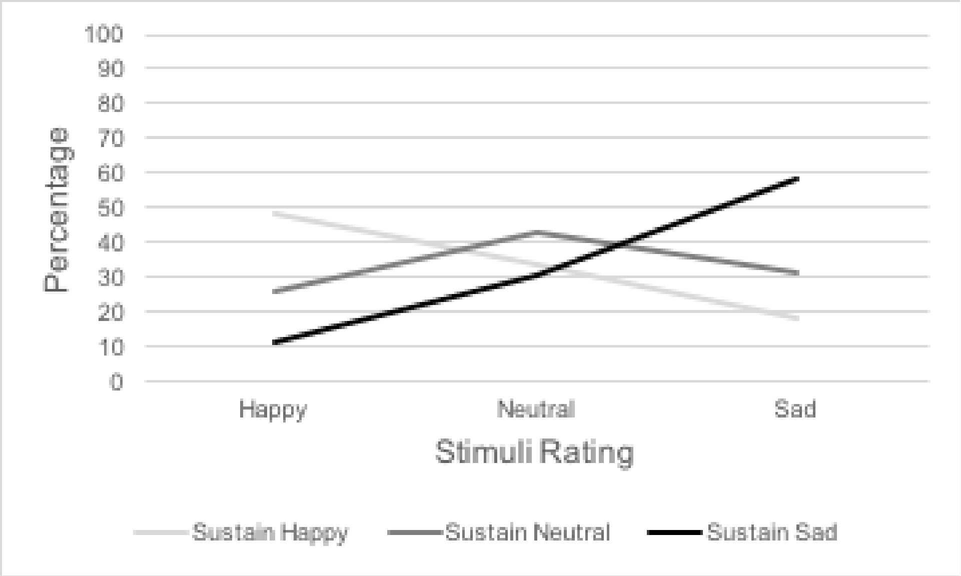


Figure 6.2. Percentage of happy (very happy and happy), neutral or sad (very sad and sad) ratings given to stimuli as a function of their sustain emotional valence.

To a lesser extent, termination also appears to be a reasonably good predictor, as seen in Figure 6.2. When termination is assigned a rating of happy, participants also tend to rate that stimuli as having an overall emotional valence of happy. Neutral ratings of termination also appear to correlate with neutral ratings of the stimuli overall. However, sad ratings of termination do not show a strong correlation with overall sad stimuli ratings. One reason for this may be that only a few terminations were assigned a very sad rating by the analyser (see Table 6.2 for a breakdown of how many stimuli were assigned a termination rating of sad/very sad, as compared to happy/very happy). This may be due to very sad terminations being defined as having a "tendency to taper off at the end" (see Table 6.1). Tapering off was something that was not often identified by the analyser when assessing termination. Reasons for this will be discussed in section 6.3.3, however for now it will suffice to say that this may be because tapering takes time, and the stimuli used in the pretest were quite short. Therefore, fewer ratings of sad/very sad terminations coupled with the tendency for participants to avoid extremes may have resulted in termination not appearing as a good predictor in sad cases.

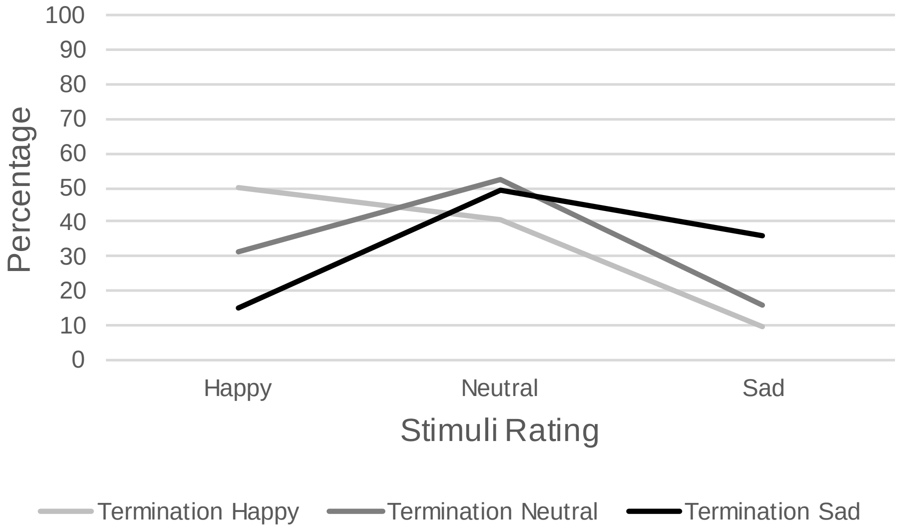


Figure 6.3. Percentage of happy (very happy and happy), neutral or sad (very sad and sad) ratings given to stimuli as a function of their termination emotional valence.

Contrast and articulation, Figure 6.4 and Figure 6.5 respectively, show less of a correlation between the ratings assigned to the TAs and participants' ratings of stimuli overall. Regardless of TA rating, most participant overall ratings were neutral, possibly with the exception of happy articulations (Figure

6.5) which did result in a relatively high proportion of overall happy ratings by participants.

Having assessed the TA ratings against participants' overall ratings of the stimuli, onset and sustain appear to be the best predictors, with termination being relatively good. Onset and sustain in general have been found to be salient components of a sound. As Kate Heidemann notes,

> [r]esearch that uses multidimensional scaling (MDS) algorithms to discover the acoustic correlates of a listeners' perceptions of timbre suggest that the most perceptually salient acoustic components of timbre are a combination of spectral centre of gravity (the dispersal and strength of frequencies above the fundamental frequency) and attack time (the nature of the onset of a sound). (Heidemann, 2016, p. 3)

The spectral centre makes up an important part of sustain, and attack time may have direct consequences on onsets. That these aspects of sound are highly salient may be one reason they have been found to be good predictors, however this does not account for termination being found to be a somewhat good predictor. In the subsequent section, I will offer two further reasons for these TAs being found to be good predictors: first, onset, sustain, and termination contain noise and, second, they generally operate on a note-to-note basis.

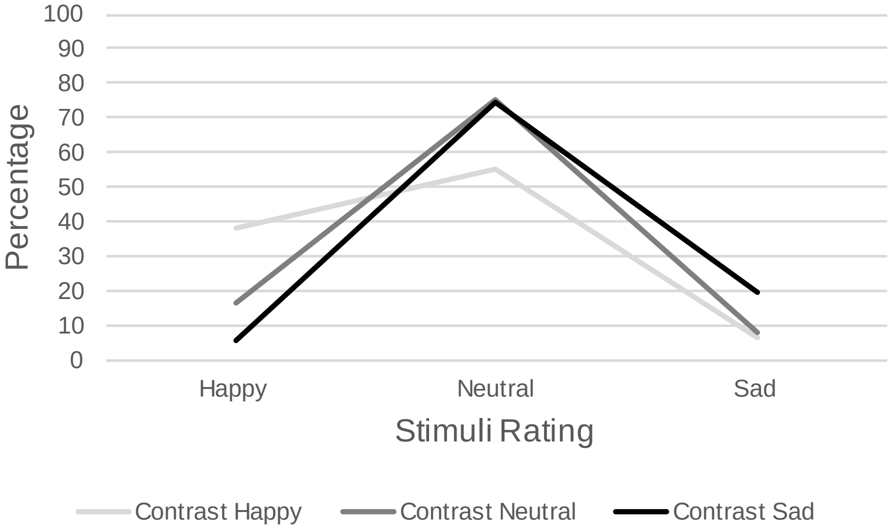


Figure 6.4. Percentage of happy (very happy and happy), neutral or sad (very sad and sad) ratings given to stimuli as a function of their contrast emotional valence.

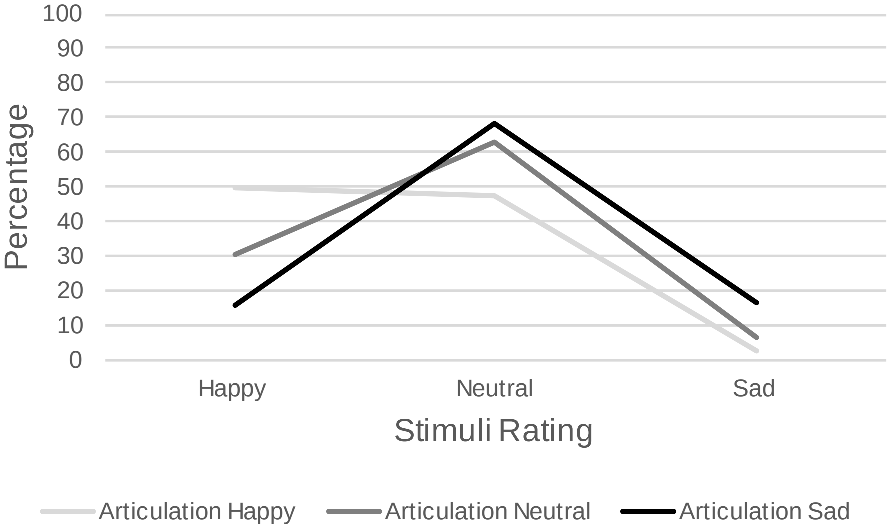


Figure 6.5. Percentage of happy (very happy and happy), neutral or sad (very sad and sad) ratings given to stimuli as a function of their articulation emotional valence.

6.3 Identifying Trends

6.3.1 Noise.

The TAs that were found to be the best predictors are also those that have the potential to contain varying amounts of noise. By noise, I refer to the use of breathy, throaty, unpitched sounds within a vocal timbre. It may be that noise is important in the perception of emotional valences, especially sad/negative emotional valences. If this is the case, then any TA that is associated with aspects of vocal timbre that have the potential to contain varying levels of noise is likely to be a good predictor.[16]

6.3.1.1 Noise potentially increasing a TA's reliability as a predictor.

Noise has been identified aurally as an important element in the TAs of onset, sustain, and termination, and this is highlighted in their descriptions (seen in Table 6.1). Noise can also be visually identified in these TAs through

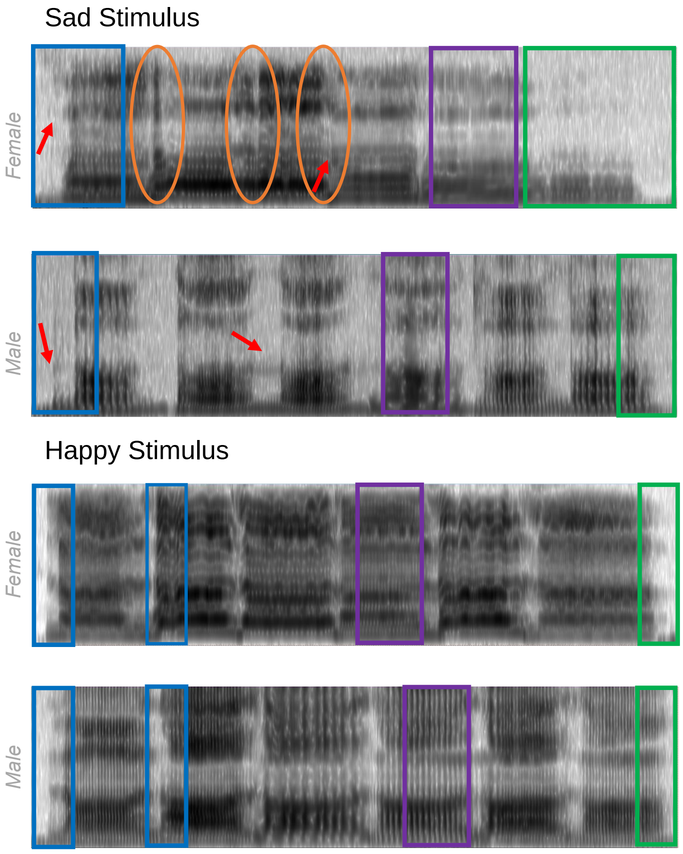
[16] Noise may have other connotations too, for example the use of breathy sounds to evoke a feeling of intimacy and, possibly, love. However, I will here focus on the connection between noise and sad/negative emotional states as it was within the dichotomy of happy/sad that the pretest was conducted. The relationship between noise and other emotive states is also worth exploring.

the use of spectrographs. The fact that noise can be identified both aurally and visually in these TAs, and that these TAs seem to be good predictors, may suggest that noise plays some part in their reliability as predictors.

Spectrographs of stimuli rated as happy and sad from the pretest can be seen in Figures 6.6 and 6.7. Throughout these examples, differences can be seen at the start of each note (the onset), through the middle of the notes (sustain), and at the ends of the notes (termination). In happy spectrographs, onsets happen quickly and clearly, and are followed by a strong sustain. Visually, happy onsets have an even spread of frequencies, beginning with all sounding harmonics present (this can be seen in the blue squares, lower panes, in Figure 6.6). Sad onsets, on the other hand, show the singer building up to the note. There is some noise present before the pitch begins. This can be seen in uneven fragments of darker patches, particularly in the upper harmonics (blue squares, upper panes, in Figure 6.6).

Another striking contrast between happy and sad vocal timbres is found in the TA of sustain. The happy spectrographs appear to show a much stronger sustain than sad ones (Figure 6.6, purple squares). This is shown by the darker patches in the happy spectrographs, which indicate stronger sounding harmonics (frequencies) above the sung note (i.e., stronger sustain).

Differences in termination can also be seen in the contrast between happy and sad spectrographs (see green squares, Figure 6.6). A closer look at the differences in happy and sad terminations, shown in Figure 6.7, reveals that sad terminations have more of a tendency to taper off, while happy terminations tend to remain quite full and strong through to the end of the note. As the sad terminations taper off, noise becomes increasingly present, which can be seen in the increasing unevenness of the sounding harmonics.



Happy **onsets** are faster and clearer. It takes more time for pitch to be observed in sad onsets. This slower speed of onset in sad stimulus allows for **noise** (in the form of breathy, throaty sounds) to occur first, followed by pitch (as shown by red arrows ↓).

Sad **terminations** take more time and have a tendency to taper off at the end. As these terminations taper off, **noise** may occur as breath and other unpitched sounds are introduced to the vocal timbre. Happy terminations occur quicker, with a clearer ending.

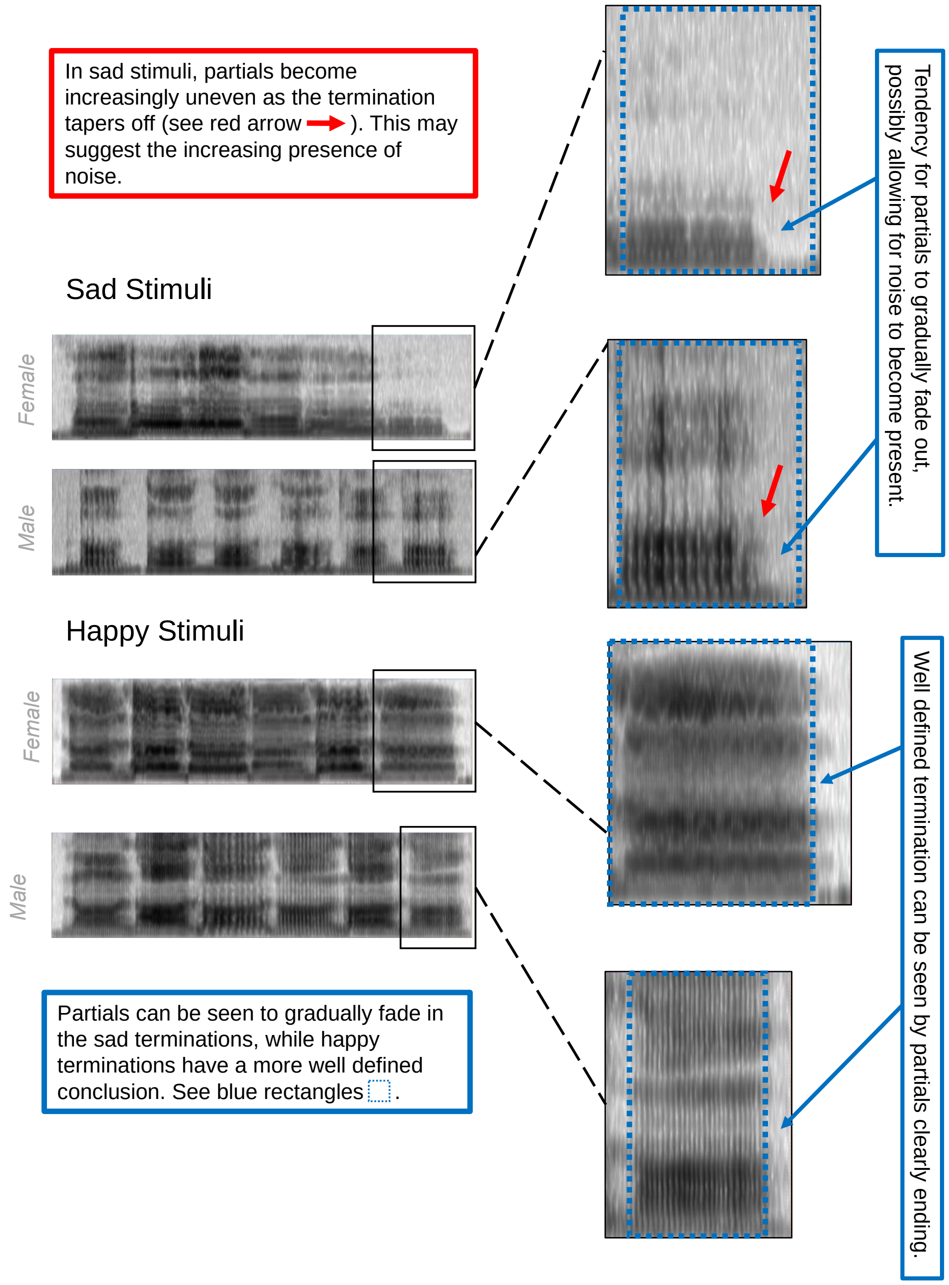
A resonant, strong **intensity** can be seen in the Happy Stimulus. The Sad Stimulus tends to have a less resonant quality shown by less density in partials. Breathy, throaty sounds may increasingly become present in these intensities allowing for more opportunity for **noise** to occur in sad intensities.

Articulation and **contrast** are difficult to distinguish in these spectrographs.

Contrast is concerned with the juxtaposition of strong, resonant timbres with less-resonant, more noisy timbres. Contrast cannot be identified in these stimuli as it takes time to be effective.

Articulation is concerned with how well each word is enunciated. In sadder articulations, there is less separation of words. While there may be some evidence of this in Sad Female stimuli (see orange circles ○), on the whole articulation is not well represented in these shorter stimuli. It is also not possible to distinguish levels of noise in articulations in these stimuli.

Figure 6.6. A comparison between TAs in sad and happy stimuli. Differences can be seen between sustain, onset and termination (see Figure 6.7 for more detailed spectrographs of termination). In particular, the levels of noise (i.e., breathy, throaty, unpitched sounds) are very different. These differences are not shown for the TAs of articulation or contrast.



In sad stimuli, partials become increasingly uneven as the termination tapers off (see red arrow ➡). This may suggest the increasing presence of noise.

Tendency for partials to gradually fade out, possibly allowing for noise to become present.

Well defined termination can be seen by partials clearly ending.

Partials can be seen to gradually fade in the sad terminations, while happy terminations have a more well defined conclusion. See blue rectangles ⬚ .

Figure 6.7. A comparison between terminations for sad and happy stimuli. Sad terminations have a tendency to taper off, allowing noise to become present in the stimuli.

The TAs of contrast and articulation, unlike onset, sustain, and termination, do not have noise as one of their key considerations. That is, noise is not as prominent in the descriptions of contrast and articulation (see Table 6.1). Noise is also not visually identifiable in the spectrographs of these TAs. In contrast to the spectrographs of onset, sustain, and termination, noise is not easily identifiable in articulation and contrast in Figure 6.6 (see orange boxes and ellipses). In the case of articulation, there is an emphasis on word delivery, rather than noise. In sadder articulations, words are not well enunciated, tending to run into one another. In happier states, articulations are clear and well separated. It is not necessary for noise to be present for this kind of unclear versus clear articulation to occur. Consequently, the spectrographs in Figures 6.6 and 6.7 do not show noise in articulations.

The spectrographs also fail to show a relationship between noise and contrast. One reason for this may be that, according to its definition, a juxtaposition between happy, resonant, belt-like timbres and sad, subdued timbre is necessary for this TA to become effective. Such a contrast may require several phrases to become established. At only six syllables long, the stimuli used in the pre-test may be too short for contrast to take effect, therefore resulting in contrast not being identifiable in the spectrographs. Longer stimuli may show contrast to be a good predictor if there is time for the juxtaposition (particularly in sadder states where noisy and clear vocal timbres are juxtaposed) to take effect. In this way, while contrast does have the potential to contain noise (as in its sadder state it juxtaposes noisy timbres with clear ones) this is not apparent in the spectrographs of the pretest stimuli.

In sum, it would seem that the TAs which have been shown to be good predictors are also those that contain noise, and this is identifiable in both their descriptions and in the spectrographs.

6.3.1.2 Noise as an inherent property of some TAs.

The fact that some TAs contain noise is not arbitrary. Rather, onset, sustain, and termination are aspects of a vocal timbre in which noise is *inherent*. That is, noise is a natural, physical and, (possibly) singer-strategic, consequence of those TAs' production. This may be because these TAs occur at certain points in the production of vocal timbre that predispose them to contain varying amounts of noise, and that in their saddest form this noise is more apparent. Onsets, for example, have the potential to be very percussive. Consider the violin bow hitting the string, the tongue releasing the airstream on the flute, and the vocal folds and mouth shaping the word and releasing the air of the sung voice. Each of these onsets are percussive in some way as each uses some level of force to initiate the sound. This force may then predispose these onsets to contain varying amounts of noise.

Noise may be similarly inherent in the TA of sustain. A note that can be prolonged needs to be sustained in a manner that maintains interest (see Chapter 3). The voice is one example of an instrument that can produce a note which can be sustained for a long duration. One may speculate that, during long notes, singers manipulate their vocal timbre to maintain interest. For this, techniques such as vibrato may be employed. However, other techniques may also be used which may not only serve to maintain interest, but also to convey emotion. Varying the level of noise present in the sustain may be one way this is achieved, for example, by the increased use of breath or the inclusion of a stylised scream.

6.3.2 Noise as an ecologically valid indicator of sadness.

The relationship between the presence of noise in onset and sustain, and their validity as predictors appears to be especially true of these TAs in their saddest forms. That is, the more noise present in these TAs, the more likely the vocal timbre is to be perceived as sad. It would seem, then, that noise may be linked to the idea of sad/negative emotional states.

One reason for the presence of noise signalling sadness in vocal timbre is that noise may play an important part in emotional perception in our everyday lives. In particular, noise may play a part in the perception of negative emotions in vocal utterances. For example, in Bradley and Lang's *The International Affective Digitized Sounds (IADS-2): Affective Ratings of Sounds and Instruction Manual,* which tested participants' perception of emotional valence of everyday sounds (Bradley & Lang, 2007), it was found that vocalisations such as screams and sobs were considered most negative. These sounds also have the most potential to contain noise in the form of breathy, throaty sounds; that is, they have some link to noisy vocalisations. Table 6.3 shows the most negative sounds according to the IADS.

Table 6.3

The International Affective Digitized Sounds (IADS) Rated as Being Most Negative

Note. IADS rated as being most negative, defined by myself as having an IADS mean rating in the pleasure dimension of less than 2—see Bradley and Lang (2007) for more information.

Sound	IADS mean score	Sound	IADS mean score
Attack 1	1.68	Attack 2	1.8
Victim	1.68	Fight 1	1.65
Male Scream	1.99	Child Abuse	1.97
Female scream 3	1.63	Female scream 2	1.93

Three of the eight sounds listed in Table 6.3 are noisy vocalisations themselves: male scream, female scream 3, female scream 2. The remaining five also have the potential to contain noisy vocalisations (for example, grunts in a fight, cries of a victim). This is interesting as the IADS tested a variety of sounds, including “car wreck”, “siren”, and “plane crash” which also have clear associations to danger and pain. However, it was those sounds which had the potential to contain noisy vocalisations that were rated as most negative by participants.

Certainly, social constructs are likely to influence this association between (noisy) vocalisation and emotion. In the Western world, we see people cry when they are sad, laugh when they are happy. Emotive sounds correspond with emotive actions with which we are all familiar. This connection between sound, action, and emotion may serve to intensify the emotive impact of such sounds. It is not just that we perceive noise as sad—it is that when we are sad, we make sounds that are noisy. Given this link can be seen between noise and emotional valence in our lives in general, it is perhaps not surprising that this should also extend to music (especially to vocal timbre) perception. In other words, noise may play a role in emotional perception of vocal timbre in the same way as, or perhaps because, it impacts on perception of emotional valence in everyday life.

Anecdotally, this connection between sound, action, and emotion was observed in the singers who recorded the stimuli for the pretest. When recording happy stimuli, singers tended to sing with an open mouth shape and the vocal production was more energetic. This produced a strong sound and emphasised the pitch over other vocal timbre characteristics. The production of sad vocal timbres however, saw singers using a more closed-off shape of the mouth and exerting less energy in the vocal production, overall it seemed to take more effort to simply produce a sad vocal timbre (perhaps in much the same way it takes more effort to engage in conversation when one is feeling sad/negative emotions). This seemed to reduce the resonance and increase the prevalence of other timbral characteristics such as breathy, throaty sounds (i.e., noise). In this way, the idea that in life we may make sounds that are noisy when we are sad may translate to vocal timbre, causing sad vocal timbre to be noisy and resulting in vocal timbres which contain noise being perceived as sad/negative by a listener.

6.3.3 Time.

Time may also affect a TA's reliability as a predictor. TAs found to be reliable predictors have the potential to operate on a note-to-note basis and therefore can be assessed and perceived at any duration, while TAs found to be less reliable take more time to be effective. Take articulation, for example. Articulation was not found to be a reliable predictor. One reason for this may be that, according to the definitions in Table 6.1, it should take time to be effective. In its sadder state, articulations are less clear, taking more time for words to be sung. In this way, articulation may be related to the musical feature of tempo. Therefore, the shortness of the stimuli in the pretest may not have allowed for articulations representative of happy and sad valences to be produced by singers.

The reduced length of the stimuli may also have made it difficult for the analyser to distinguish where termination ends and articulation begins. According to the definition, a sad articulation may include more continuity of sound. This potential for sad articulations to be fluid, with sounds running into each other rather than being discreetly defined, may have made it difficult for the analyser to rate the emotional valence of articulation independently of that of termination. Additionally, in sad stimuli, participants may also have been more influenced by termination, with the effect of articulation being overshadowed. This would explain why happy articulations (where sung *la's* were clearly separated) seem to correlate more clearly with participants' ratings than sad articulations (where sung *la's* were not clearly separated). This can be seen in Figure 6.5.

Contrast may also require time to be effective. According to its definition, a happy contrast consists of *belting*—a vocal timbre that sounds at the edge of its range, but not so far that it compromises other musical elements. A sad contrast on the other hand consists of building timbre—building anticipation/emotional interest by contrasting subdued timbres with fuller timbres. Due to the reliance on this juxtaposition of subdued and belting timbres, it seems likely that more time is required for contrast to impact on listener perception.

The question of time may also explain why termination was found to be only a somewhat good predictor when compared to onset and sustain. Onset and sustain are both defined in terms of their function within a single note. Termination, however, has a level of dependency on subsequent notes. In its sadder form, termination has a tendency to taper off. This suggests a reliance on time as time is needed in order for the tapering off to be effective. It may be, then, that termination's reliability as a predictor would increase in more ecologically valid contexts.

Onset and sustain may have been found to be such reliable predictors not only because they contain noise, but also because they operate on a note to note basis and, thus, do not require too much time to be effective.

6.4 Some Caveats

There are several key points that need to be considered in relation to the TAs and their implications for vocal timbre analysis. First, as noted at the outset of section 6.2, the testing of the TAs was not the main purpose of the pretest. Therefore, the results of the comparisons of the attributes to overall stimuli ratings can only be considered preliminary at this stage. As such, the discussion presented here is somewhat speculative in nature. While it is possible to theorise about the impact of the TAs on emotional perception from the pretest findings, further, more targeted research, is required to validate these conclusions. Nonetheless, these results do strongly suggest that TAs have the potential to aid in vocal timbre analysis by assisting in the classification of emotional valence in vocal timbre.

Second, the ratings of the TAs were completed by the analyser (i.e., myself) alone. In further research, the TAs should be rated by a panel of expert judges to ensure that the interpretation of emotional valence in each TA is not idiosyncratic (i.e., specific to the analyser). Nevertheless, we may assume that the ratings applied to the TAs here are quite intersubjective as, anecdotally, they were agreed upon by the singers emulating the vocal timbre tokens, and also given the subsequent correlations with the overall score given by participants.

Third, when discussing the usefulness of the TAs in defining emotional valence in vocal timbre, one must assess how easily and accurately they are able to be identified as discrete attributes. Some TAs (like onset) lend themselves very easily to being identified as a discrete attribute. It is easy to discern where the onset begins and (usually) where it ends. Other TAs are less clear. For example, termination, where one cannot always easily identify the beginning. The TAs may require further refining so that they may be easily identified.

6.5 Conclusion

Through a process of comparing the emotional valences of certain acoustic features of timbre with emotional valence ratings of the participants in the pretest, onset, sustain and termination have been identified as the best predictors of emotional valence in vocal timbre. It is possible that the reason these TAs align with timbral valence is that they contain varying levels of noise, and noise appears to be a feature of sad/negative vocalisations more generally. However, because it was not the goal of the pretest to test the predictability of these TAs, their reliability requires further investigation. How each TA individually affects emotional perception also requires further investigation: Is it possible that within a vocal timbre, some TAs are at times overshadowed by others? That is, do some TAs impact on emotional perception of a vocal timbre more than others? Further developing the ideas presented in this chapter could have consequences for future vocal timbre analysis as it may allow the analyser to easily and accurately identify emotional valence in the vocal timbre and allow us to better understand music perception in general. Nevertheless, the evidence presented thus far seems sufficient for at least some of the TAs (namely onset, sustain, and termination) to be used to assist in the identification of emotional valence in vocal timbre.

Part III: Methodology and Application

Readers must recognise one problem from the start. It is the hardest thing imaginable to find a descriptive and analytical language that does justice to any newly emerged form of musical expression. (Bowen, 1970, p. 33)

Singing is like loving somebody…It’s a supreme emotional and physical experience. (Janis Joplin, cited in Hughes, 2018, February 6, para. 3)

7 Methodology: Outline of a New Analytical Technique for Vocal Timbre in Popular Vocal Song

7.1 Introduction

This chapter describes a new approach to vocal timbre analysis which aims to fill the gap of analytical methods for vocal timbre. To circumvent the difficulties associated with analysing a musical feature that is not traditionally notated, which is a contributing factor in it having been so seldom analysed in the past, the approach presented here is multilayered. That is to say, in this approach, the song is examined at a variety of levels. The recording always serves as the basis of analysis. The emotional valence of the vocal timbre (which may change from moment to moment) can be identified using the Vocal Timbre Features, which is a system that I have developed to aid in the identification of certain acoustic features within a vocal timbre that may assist in categorising the emotional emotional valence of a vocal timbre (as has been argued in Chapter 6, and as will be discussed in more detail in section 7.3). A set of signs and symbols have been assigned to these Features, allowing them to act both as a system of identifying emotional valence and as a method for succinctly annotating vocal timbre. Observations made by aurally detecting and annotating the Vocal Timbre Features can be confirmed visually through spectrographs. The emotions potentially indicated by vocal timbre interact with the messages conveyed by lyrics in complex ways—the lyrics potentially helping the listener to interpret the emotions indicated through vocal timbre, and the vocal timbre potentially influencing the interpretation of the lyrics.

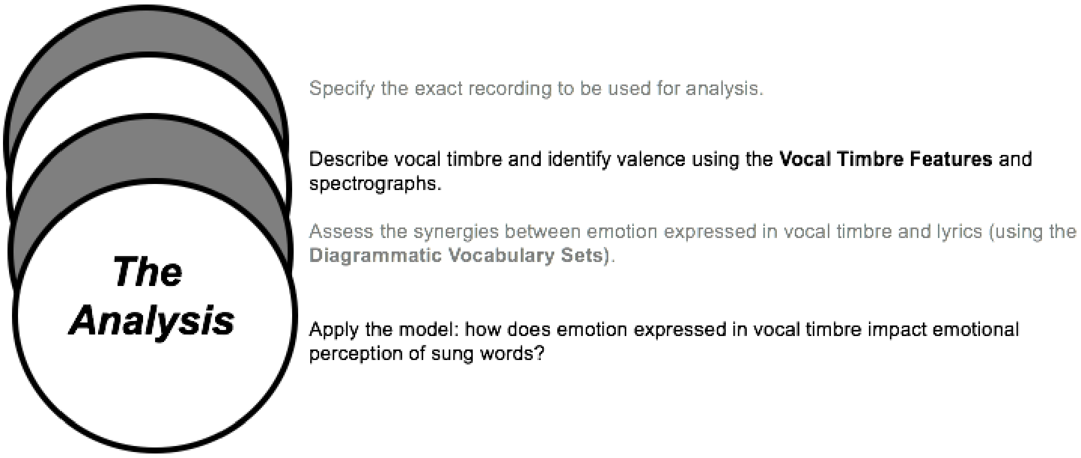
While this chapter will unpack each of these layers in turn, it is useful to begin with a structural overview of the technique. Figure 7.1 provides a visual representation of the steps present in this analytical technique.

The first step is to select the recording to be analysed. It is important to specify the exact recording because of the possibility for every performance and associated recording to be different. This is discussed in more detail in section 7.2.

The next step is to describe the vocal timbre heard in the selected recording. This is achieved by means of the Vocal Timbre Features and their corresponding symbols, and the use of spectrographs (to visualise the vocal timbre and confirm observations made regarding the Vocal Timbre Features). Both Vocal Timbre Features and spectrographs are documented "on the page" (i.e., they can be reproduced in the written analysis, as in Chapter 8), and as part of the video descriptions of vocal timbre which syncs both audio (i.e., recording) and visuals (i.e., Vocal Timbre Features and spectrographs, among other things, see 7.3.3 for a more detailed discussion). Describing vocal timbre in this way not only lends clarity to the analysis, but it can also assist in identifying emotional valence. These tools are discussed in more detail in section 7.3.

Following the description process, the potential synergies between emotion conveyed in vocal timbre and emotion conveyed in words are assessed. This is discussed in more detail in section 7.4.

One aspect which permeates throughout the analysis is the application of the model. That is, to explicate vocal timbre on the basis of its emotional indicators and draw inferences as to how this affects the interpretation of the lyrics. This step is implemented more or less concurrently with the previous processes as it is the model which allows one to explore the synergies between emotionally valenced vocal timbres and emotional messages in lyrics.



Specify the exact recording to be used for analysis.

Describe vocal timbre and identify valence using the **Vocal Timbre Features** and spectrographs.

Assess the synergies between emotion expressed in vocal timbre and lyrics (using the **Diagrammatic Vocabulary Sets**).

Apply the model: how does emotion expressed in vocal timbre impact emotional perception of sung words?

Figure 7.1. The Analytical Structure. Visualisation of the components which make up the analytical approach to vocal timbre proposed in this thesis. Each of these components contribute equally to the analysis.

7.2 Analytical Scope

The scope of music to which this analytical approach could be applied has been touched on in Chapter 1 of this thesis. In brief, popular music where the vocal line plays a key role in conveying meaning and emotion (hence forth called “popular vocal songs”) was identified as most appropriate for the kind of analysis proposed here. There are two key attributes of such songs which make them ideal texts for analysis. First, in such songs, the lyrical content is easily intelligible. This is because it is the lyrics that are typically used to convey the message of the song. Therefore, one must be able to understand the words being sung. Second, as will be seen in the following paragraphs, as the lyrics are important in supplying cues about the songs meaning, the vocal line is usually prominent in the mix. This serves to emphasise the vocal line and, by extension, vocal timbre and lyrics. In this way, these kinds of popular vocal songs provide a good context in which to examine the vocal line.

Specifying this scope is important as choosing the right context in which to study vocal timbre in the way I propose is essential to the success of the analysis. In his 2007 book *Listening and Voice,* Don Ihde wrote that “[t]hings or objects appear only as they are essentially situated in a field” (Ihde, 2007, p. 38). In other words, how we perceive the importance of certain “things” (such as musical features) depends on the context in which they are encountered. Ihde illustrates this idea using the following scenario:

> if a professor asks her class, is black a color? she will be likely to receive a number of negative answers. As reasons for their answers the

> students will recite what they have learned concerning color from the sciences, perhaps claiming that a "real" color is defined in terms of wavelengths of light. But if the professor's question is, What color is that? while pointing to the blackboard, the overwhelming answer will be, simply, black. (Ihde, 2007, p. 27)

Whether black is or is not a colour depends on the way the question is asked. That is to say, the question's context plays a key role in shaping the answer to the professor's questions relating to colour perception.

Likewise, the musical context also plays a key role in the way one approaches vocal timbre analysis. A piece may be created such that certain musical features play a central role in the creation and perception of emotional meaning. Popular vocal songs that are of interest to this study are those where the vocal line, and thus vocal timbre and lyrics, are emphasised (i.e., they "appear" foregrounded against other musical features).

Various strategies can be used to draw attention to a vocal line through mixing and production. Vocals in popular music are typically mixed in a way that helps to foreground and draw attention to the voice or to make the voice stand out. It is not the goal here to engage in an in-depth analysis of vocal placement in popular vocal songs, but rather to demonstrate that the placement of vocals in general foregrounds the voice in such music (and in the case studies in Chapter 8). Allan F. Moore provides a model through which instrument placement might be assessed: the soundbox. The soundbox "provides a way of conceptualizing the *textual space* that a recording inhabits, by enabling us to literally hear recordings taking space" (Moore, 2012, p. 30).

Since the 1970s, instruments have tended to be positioned in a similar way across recordings:

> Through the latter 1960s, producers and engineers gradually came to adopt a normative positioning of sound-sources within the soundbox, a positioning that tends to remain stable throughout the track. I call this the diagonal mix in recognition of its determining factor. The operative elements in such positioning are a lead voice, a snare drum and the harmonic bass (normally a bass guitar), which are situated centrally on a (very) slight diagonal. The resultant diagonal mix tends to operate normatively for all genres from the early 1970s to the present. Two other possibilities tend to dominate, both of which can occasionally be found to this day. One of these was the cluster mix, whereby sounds were grouped rather tightly within the available space. The other possibility was the triangular mix. If the diagonal mix observes a line through the lead voice, snare drum and bass, the triangular mix has these in three different positions, with two to one side of the mix and one to the other. (Moore, 2012, p. 32)

In all of the common mixes, the vocals sit at or near the centre, thus foregrounding vocal timbre and lyrics. Certainly, this mix can be varied (Moore gives examples of this), however, for a large part of popular vocal song the placement of the vocals at or near the centre is normative. Examining songs that foreground the vocal line in this way, and that use lyrics to “tell a story”, seems to be a logical starting place for examining the role of vocal timbre and the way it may contribute to the emotional meaning of a song.

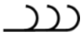
Since vocal timbre is not usually notated, its analysis relies on recordings. Furthermore, given the potential for timbre to vary substantially from one performance to another, an analysis such as the one proposed here will only be able to account for the specific timbre used during a specific performance of a given song. For this reason, the exact recording to be used must be specified.

This approach is in line with a growing trend to use recordings in musical enquiry and analysis (see Chapter 4 for more discussion on phonomusicology). As Cottrell puts it, the study of "recorded sound is now an important component for a number of subdisciplines of music studies" (Cottrell, 2010, p. 16). Popular music studies, and the study of vocal timbre analysis in relation to popular songs, is one area that relies on recordings. It is for this reason that, in the analytical approach detailed here, the recording is taken as the primary analytical text.

7.3 Describing Vocal Timbre and Identifying Valence

Two techniques may assist in the process of describing vocal timbre: the Vocal Timbre Features and spectrographs. This section will look at each of these tools in turn and describe their application to the analysis. At times, general terms such as noise are used in place of more technical ones, such as "noise" and "creak". This is a deliberate decision as it is the goal of the technique presented here to be easily used by researchers from a variety of backgrounds.

7.3.1 Vocal timbre features.

The Vocal Timbre Features constitute a classification system to assist in identifying and describing certain acoustic cues within a vocal timbre. The identification of these cues may also be used to categorize emotional valence within a vocal timbre. Each of the features of the Vocal Timbre Features has been assigned a unique symbol which can be used in place of the verbal description (for example, a breathy sustain is symbolised as ). The classification system was created by drawing on my own findings (see Chapter 6) as well as that of other scholars who have attempted to describe vocal timbre (in particular, I have drawn on Van Leeuwen (1999), Heidemann (2016), and the Jo Estill method for singing (Mc Donald Kilmek, Obert, & Steinhauer, 2005)). An important component of the system consists in the identification of certain acoustic features within three main components of a vocal sound: onset, sustain (like steady state or sustain, as it is sometimes referred to in other literature, see Chapter 3), and termination.

In Chapter 6, it was shown that the form of certain acoustic traits (called Timbral Attributes, or TAs for short) correlated with how happy or sad a vocal timbre was perceived to be. These TAs and their definitions can be found in Table 6.1. The TAs of onset, sustain, and termination were identified as the best attributes to assess emotion. Because of this, and because these components have been identified as important in other literature too (e.g., Erickson, 1975; Heidemann, 2016), onset, sustain, and termination were used as the basis for the Vocal Timbre Features, forming the three main components developed here as tools to describe vocal timbre and, potentially, aid in classifying emotional valence through the Vocal Timbre Features.

Being able to succinctly describe acoustic features of a vocal timbre has two benefits:

- *It helps to achieve clarity and efficiency in analysis.*

  The ability to describe a musical feature makes its analysis quicker and clearer. For example, being able to say "that pitch is C" or "that note is a crotchet" affords clarity to the discussion of pitch and rhythm. This clarity is achievable because there exists a predetermined system of notating such elements as pitch and rhythm. Having a system to describe aspects of a vocal timbre through a set of discrete symbols would, then, also lend vocal timbre analysis a level of clarity which may otherwise not be available.

  While this section is concerned with the Vocal Timbre Features, it should be noted that spectrographs (i.e. a visual representation of sounding harmonics present within a sound) are also used in the analytical approach developed here. The role of spectrographs is discussed in section 7.3.2. Together, the Vocal Timbre Features and spectrographs will be used to:

1. confirm visually what has been heard aurally;
2. demonstrate musical landmarks and how these may contribute to the aural experience of the popular vocal song; and
3. complement the listening of musical examples and enhance the reader's understanding of concepts presented within the analysis.

- *It may help to identify emotional valence.*

  Identifying and describing the form of certain acoustic features within a vocal timbre has the potential to help in categorising the emotional valence of that vocal timbre in two ways. First, it allows one to assess the amount of noise, and other non-pitched sounds, present in a vocal timbre. As discussed in previous chapters (and as will be elaborated in the following paragraphs), it has been hypothesised that the more noise (i.e., throaty, non-pitched sounds) present within a vocal timbre, the more likely it may be perceived as sad/negative (see Chapter 6). Other associations may also exist between the form of certain vocal features and emotional state (e.g., the sound of the breath may evoke a sense of intimacy, as discussed as follows). While further empirical investigation is certainly necessary to determine the relationship between noise and other non-pitched sounds, it is possible that identifying the level of noise in a vocal timbre may assist in classifying emotional valence.

Second, describing vocal timbre often involves describing how certain timbres are produced (as will be seen in the following paragraph). There is evidence to suggest that, upon hearing a vocal sound, a listener may replicate within themselves the process used to produce such a sound.

For example, upon hearing a scream, a listener may recall the experience of producing this sound, in much the same way one would if actually screaming. There are several theories which suggest that we may understand emotions through such a system of embodiment (see Chapter 3, section 3.3.3, Chapter 4, section 4.4.4, and Chapter 9, section 9.2.1 for a more detailed discussion on ideas relating to embodiment). The idea that vocal timbre may telegraph "the interior state of a moving body, presenting the listener with blueprints for ways of being and feeling" (Heidemann, 2016, p. 1) is useful for classifying emotional valence as it suggests that, if one can identify the process by which a singer produces a sound, then one may also be able to speculate about the impact of that sound on listeners. In these ways, the Vocal Timbre Features can aid in identification of emotional valence—something which is important for the analytical approach outlined in this chapter.

It is important to note here that, in using the Vocal Timbre Features to identify emotional valence, the importance of context cannot be overstated. For example, a tense vocal timbre may indicate stress, but it might also indicate excitement. To make a judgment about which, the analyser needs to assess not only the acoustic features of the vocal timbre, but consider these in their wider context too (what has happened in the vocal timbre previously? what is happening in the lyrics?). How context may impact emotional judgments will be discussed in relation to the specific Vocal Timbre Features in sections 7.3.1.1–7.3.1.3.

The remainder of this section will present the Vocal Timbre Features in detail. As discussed above, the Vocal Timbre Features are broken up into three main components: onset, sustain, and termination. Onset and termination are measured along one dimension each, and sustain is measured along six dimensions. This can be seen in Figure 7.2, where the Vocal Timbre Features and their corresponding labels are shown. I have chosen to measure sustain along a number of dimensions as it is often the longest part of the note and is often comprised of “a mixture of different features” (van Leeuwen, 1999, p. 129). Measuring sustain along six different dimensions allows the analyser to better describe the complex nature of the vocal timbre heard.

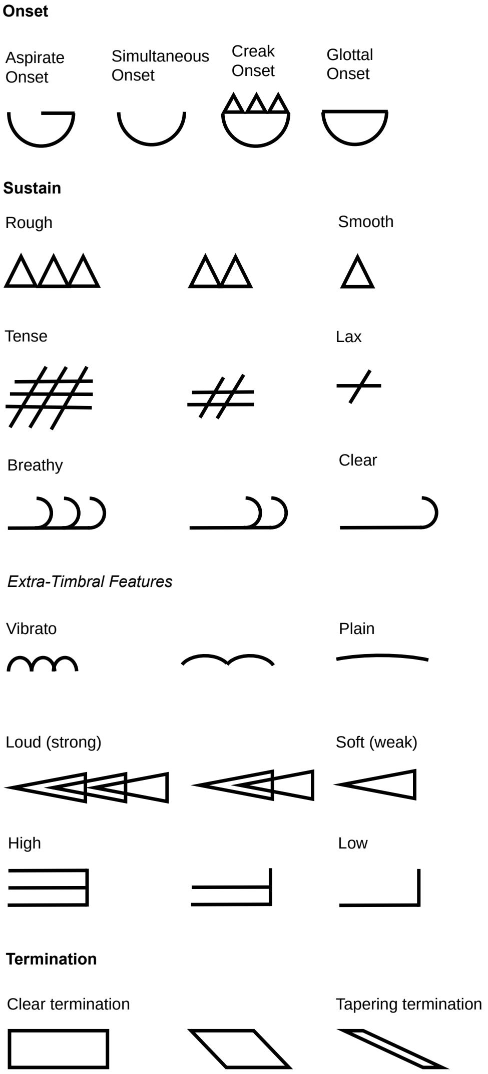


Figure 7.2. List of Vocal Timbre Features (and their corresponding graphic representations) which can be used to describe vocal timbre and identify emotional valence.

7.3.1.1 Onset.

Onsets are characteristic features of musical sounds in general (see Chapters 3 and 6 for a more detailed discussion). Jo Estill (who developed a singing method that centres on the control of specific vocal tract structures (see McDonald Kilmek et al., 2005, pp. 2–4), defines three different classes of onset in vocal timbre specifically: glottal onset, aspirate onset, and simultaneous onset.[17] I propose adding one more kind of onset to this list based on Heidemann's 2016 paper "A System for Describing Vocal Timbre in Popular Song"—the creak onset. Figure 7.3 shows these onsets represented graphically. These onsets are defined as:

[17] Some of these labels resemble those used by phoneticians to describe phonemes, the sounds of a language. In particular, *aspirate* and *glottal* are terms used to classify phonemes. While there may be some overlap between the terms as used in this thesis and as used in phonetics (after all, singing may be considered a stylised form of speaking, and therefore may draw on many of the same processes of vocal production), this does not mean that they are synonymous or that the timbral feature will be determined exclusively by the properties of the phoneme being sung.

Aspiration is not an obligatory feature of English phonemes but it occurs spontaneously for certain phonemes in word initial position. The aspirate onsets described here are probably most related to the production of voiceless phonemes, since these are produced by breath passing through open vocal folds. Nevertheless, it is possible for a vocal timbre onset to be characterised as aspirate even if the sung phoneme does not require aspiration in English. At any rate, the potential overlap between the phonetic requirements and the timbral features will be taken into account in the analyses.

- Aspirate Onset: the breath passes through the vocal folds before they begin to vibrate, giving a breathy sound. (Chate, n.d., para. 7)
- Simultaneous Onset: breath and vibration occur at the vocal folds at the same time giving a balanced tone. (Chate, n.d., para. 7)
- Creak Onset: the vocal folds open and close abruptly while the breath passes through. (Heidemann, 2016, p. 6)
- Glottal Onset: the vocal folds begin vibrating before the breath arrives, giving a hard vocal attack. (Chate, n.d., para. 7)

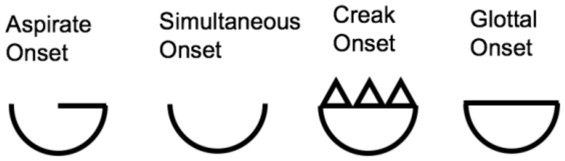


Figure 7.3. Graphic representation of the four different onsets used in this analytical technique: aspirate, simultaneous, creak , and glottal.

Based on evidence from Chapters 3 and 6, I suggest that these four types of onsets may be associated with different emotional responses.

Glottal and creak onsets

Glottal onsets are characterised as very percussive. This is because they are created by air "suddenly and forcefully escaping through a previously tightly closed glottis" (Heidemann, 2016, p. 6). Creak onsets can also be percussive, however they tend to be characterised by increased levels of roughness. This is due to the vocal folds rapidly opening and closing as breath passes through.

In both glottal and creak onsets, there is the opportunity for rough, nonpitched, noisy sounds to become present. As discussed in Chapter 6 and also noted in the preceding paragraphs, increased roughness seems to correlate with increased perceptions of sad/negative emotions. As such, glottal and creak onsets may tend to be associated with feelings of sadness, and potentially other negative emotions (e.g., weariness, pain, sadness, fear), in the listener. These onsets, as well as being noisy, also tend to be very audible (it is difficult to produce a soft glottal onset, although creak onsets may have more dynamic variety). Therefore, glottal and, at times, creak onsets may also be associated with feelings of *externalised* negative emotions. That is, emotions which one typically expresses loudly and openly, either intentionally or involuntarily. For example, a scream of fear, or a yell of pain.

Aspirate onsets

Aspirate onsets, on the other hand, are softer. They occur when breath passes through the vocal folds before they begin to vibrate (this may occure both while exhaling or inhaling). This creates a breathy effect which is noisy and unpitched, but not necessarily rough. Therefore, aspirate onsets are like glottal/creak onsets in that they have the potential to be sad/negative. However they may signify a different kind of sad/negative emotion compared to creak/glottal. The noisy but *soft* nature of the aspirate onset may signify more internalised negative/sad emotional states, for example the sound of a quiet sob or the airy quality of a frightened voice. Aspirate onsets may also have the potential to evoke a sense of intimacy or closeness as, to hear such soft nuances, one needs to be in close proximity to the singer (see discussion on breathy/clear intensities in section 7.3.1.2 for more detail on potential connotations).

Simultaneous onsets

Simultaneous onsets involve the breath and the vibration of the vocal folds occurring at the same time. This creates a clear onset where pitch begins without delay. Such onsets may be more likely to be associated with neutral and happy emotions. Consider the clear onsets of laughter, or the deliberate articulation of the prime minister's speech. In this way, simultaneous onsets may be associated with neutral and happy emotive states.

7.3.1.2 Sustain.

Sustain forms the middle, and often the longest, part of the note. Because of its longer duration, there is ample opportunity for multiple features to present themselves in a vocal timbre's sustain. Sustain may also be highly varied as these features have the potential to undergo extreme changes within a short period of time. In the remainder of this section, I have drawn on six categories of voice qualities outlined in van Leeuwen's discussion in *Speech, Music, Sound* (1999, pp. 129–141). I have chosen to use these features as they seem to be some of the most salient features associated with sustain. Using my own findings (from Chapters 5 and 6, especially drawing on the role of noise), these features have been expanded on such that they may be used to identify emotional valence in vocal timbre. The work of Heidemann (2016), Lacasse (2010), and Poyatos (2002) have also been incorporated into the features listed in the remainder of this section.

Three features discussed in the sustain have been included under the heading of extra-timbral: vibrato/plane, loud/soft, and high/low. These features have been included because they have the potential to impact on a vocal timbre. In the definition of vocal timbre given in section 2.2, frequency was acknowledged as impacting on timbre (Rossing, 1990, p. 30). Frequency may be manipulated within a vocal line through both register (high/low) and pitch oscillation (vibrato/plain). As will be seen in the discussion below, some Vocal Timbre Features are more easily produced at particular dynamic levels. For example, a softer dynamic is necessary for a breathy vocal timbre. Additionally, these features do not require much time to become apparent, rather they can be assessed at any point in the vocal line. This is unlike other features such as tempo, which may impact timbre but which also happen over an extended period of time. Therefore, while examining features such as tempo would be interesting, at present the extra-timbral features explored in this section have been limited to the three named above as these are often the most immediately apparent.

Rough/Smooth

A rough voice "is one in which we can hear other things besides the tone of the voice itself" (van Leeuwen, 1999, p. 131). "Much of the effect of ‘roughness’ comes from the aperiodic vibration of the vocal cords which causes noise in the spectrum” (Laver, 1980, p. 128 as cited in van Leeuwen, 1999, p. 132). This is also sometimes called *vocal fry* or *growl*, and is created by tensing the vocal folds and holding them tightly together (Heidemann, 2016, p. 6). This noisy quality present in roughness may signify negative emotions (see Chapter 6). For example, a scream is rough and noisy, and a scream may be considered negative. In this way, roughness and noise in vocal timbre may signify negative emotional states.

Breathy/clear

Breathiness can occur when “extraneous sound mixes in with the tone of the voice itself” (van Leeuwen, 1999, p. 133). It is “characterised by the sound of air leaking through an incompletely closed glottis” (Heidemann, 2016, p. 5). When a breathy sound is produced by vocal folds which are low in tension, the resulting sound is soft. When it is produced by vocal folds which are high in tension, the sound has more of a “hissing or grainy” quality (Heidemann, p. 5). The breathy voice may represent a number of emotive states. The first is closeness as the breathy voice is “always also soft, and fervently associated with intimacy” (van Leeuwen, 1999, p. 133). For example, a whisper is a breathy sound, and to hear a whisper one needs to be in close proximity to the speaker. From a para-linguistic perspective, Poyatos has identified the breathy voice in terms of a whisper as having the potential to express a sense of anticipation, “fear, surprise, expectancy, or sheer terror” (Poyatos, 1993, p. 202). Consider, for instance, the ragged whispering heard in horror films as the victim telephones for help. Breathiness may also be associated with the “uncontrollable nonverbal expression of sexual arousal” (Poyatos, 2002, p. 31). For example, the opening lines of Marilyn Monroe’s performance of “Happy Birthday Mr. President” (Missmalevolent, 2007). I would also like to suggest that breathiness in the voice may indicate vulnerability. For example, a sobbed utterance or the ragged breathy quality of a fearful voice. The meaning of breath in the voice (intimacy, vulnerability, fear) depends heavily on context. The emotions these utterances signify are quite universal. Everyone has, at one point, felt fear or sadness, or had to listen closely to another person’s whisper. In this way, breathiness also signifies emotional states that are close to one’s own real life experience, and therefore has the potential to be highly evocative.

Tense/lax

To sound tense one constricts the muscles in the body, particularly the throat; to sound lax one relaxes these muscles. As van Leeuwen puts it, “[t]he sound that results from tension not only *is* tense, it also *means* ‘tense’—and *makes* tense” (van Leeuwen, 1999, p. 131). That is, when we hear a tense voice, we may not only extrapolate information about a speaker’s physical state, but we may also gain a sense of their emotional state. For example, the speaker may be nervous. These cues may also influence our own emotional state. If the speaker is nervous, the listener too may begin to feel nervous.

There are many situations in which tension may be present in the voice. Consider the increasing tension in the voice as one becomes more and more frustrated with a bank teller. Or the tension in the waiter’s voice as they explain to the guest of honour that the custom-made birthday cake has been dropped. Each of these situations will likely evoke a very different response from the listener. Therefore, tension also relies on context.

Loud (strong)/soft (weak)

This feature is related to distance and power, both physical and social (van Leeuwen, 1999, p. 133). Loudness is related to strength in the sense that louder sounds are stronger in volume and therefore can signify that a listener is in closer proximity to a sound source. Loudness may also be associated with more powerful sounds and thus be an indicator of importance. Soft sounds, on the other hand, may signify distance between the sound's source and listener. Softer sounds may also be associated with less power and importance as these sounds may play a secondary role to loud sounds. Consider the softer volume of backup singers in a band, or of the sotto voce of a pit orchestra during a dialogue scene in a musical. In this way, soft/loud sounds can signify power and proximity (strong), as well as physical and social distance (weak).

High/low

This refers to register and is connected to the changing location of sympathetic vibrations within the body (Heidemann, 2016, p. 8). The use of high/low registers may be associated with ideas of dominance and assertiveness. Van Leeuwen states that men who mean to assert their dominance may speak in a higher register, while women who mean to do the same may speak in a lower register (i.e., both will use the opposite register to their normal speaking voice). High/low singing also has an impact on vocal timbre. Falsetto may be used when men and women (for women falsetto is sometimes called “head voice”) sing high. This falsetto results in a very different timbral quality of the voice, when compared to singing in the lower register where intimate/soft sounds are more easily achieved. Singing in falsetto and whistle registers may also evoke ideas of effort, as to produce these sounds one must focus vibrations in the top of the head/sinuses (especially when singing in the whistle register) (Heidemann, 2016, p. 8).

Vibrato/plain

In general, vibrato is “a family of tonal effects in music that depends on periodic vibrations of one or more characteristics in the sound wave” (Rossing, 1990, p. 134). Both vibrato and non-vibrato sounds may have the potential to evoke emotional responses in listeners. As van Leeuwen puts it:

> vibrato literally “means what it is”. The vibrating sound literally and figuratively trembles. What makes us tremble? Emotions. (van Leeuwen, 1999, p. 134)

However, “*[n]ot* trembling, sounding plain and unmoved can also acquire a variety of contextually specific meanings” (van Leeuwen, 1999, p. 135). So, vibrato and nonvibrato sounds may have the potential to evoke a range of responses. On the one hand, vibrato may signify love, tension, fear, and anticipation while nonvibrato may signify steadiness, an unmoving attitude, resolution, or acceptance (van Leeuwen, 1999, pp. 134 – 135). The specific response evoked by vibrato depends on the context in which it is present.

7.3.1.3 Termination.

Termination is measured across one dimension: *Clear/tapered*. A clear termination is one that has a strong ending; a tapering termination is one in which the sound slowly fades, allowing for other elements (such as noise) to become present. Clear terminations tend to be associated with happier emotional states, while tapered terminations tend to be associated with sadder ones (see Chapter 6 for a more in-depth explanation).

7.3.2 Spectrographs

As mentioned above, spectrographs will play a role in this analytical approach. Spectrographs are used to clarify observations made through the application of the Vocal Timbre Features. For example, if, through the process of annotating a vocal timbre using the Vocal Timbre Features, one was to aurally identify a series of breathy onsets, it would be possible to generate a spectrograph of the onsets in question to confirm these claims. This is because breathy onsets may display a certain pattern of harmonics. Visualising this pattern through spectrographs, then, can reinforce observations made in the Vocal Timbre Features.

7.3.3 Using the vocal timbre features and spectrographs in audiovisual descriptions of vocal timbre.

The audiovisual description of vocal timbre is achieved through incorporating spectrographs and Vocal Timbre Features (as well as presenting lyrics and marking places of audible breath) into one cohesive presentation (to be illustrated with the case studies presented in the next chapter). This allows the different features of a vocal timbre to be described as they are heard. Therefore, the association between the features present within the vocal timbre and the corresponding graphic representation is clearer. The major benefit of such an approach is that it helps to improve the clarity of the description and retains emphasis on the aural sensation of the vocal timbre.

Currently this is a very time-consuming process. The audiovisual descriptions presented in the next chapter have been done by hand using ScreenFlow (Telestream, 2016) software. Streamlining this process through the use of computer programs would make the audiovisual description a much more practical inclusion for this analytical technique. While developing such programs is beyond the scope of this thesis, the audiovisual description used in the next chapter demonstrates what such an analysis would look like.

7.4 Exploring the Relationship Between Vocal Timbre and Lyrics

The previous section (7.3) was concerned with how one may describe a vocal timbre and classify its emotional valence. The present section will outline how one may explore the synergies between the emotion identified in vocal timbre and the emotion conveyed by the lyrics.

Finding an accessible "vocabulary to describe sound events, structures and spaces" is central to the success of any analysis that explores music in the acousmatic (i.e., primarily aural) sense (Blackburn, 2009, p. 1). Without these accessible vocabulary sets the analysis can become difficult to control. Therefore, in much the same way that the analysis of other musical features such as harmony and melody achieve efficiency and clarity by being grounded within the parameter of pitch, vocal timbre analysis can achieve efficiency and clarity by being grounded in linguistic parameters. These linguistic parameters, once established, can then be used to explore the synergies between the emotion conveyed through vocal timbre and that conveyed in lyrics.

To this end, I propose adopting the concept of diagrammatic vocabulary sets to the analysis of vocal timbre. Diagrammatic vocabulary sets is an approach developed as part of Denis Smalley's spectromorphological analyses of electroacoustic music. Spectromorphology is a method of electroacoustic music analysis (Smalley, 1986, 1997) that seeks new ways to "discuss musical experiences, to describe the features we hear and explain how they work in the context of the music" (Smalley, 1997, p. 107). This, too, is the goal of my approach to vocal timbre analysis.

Here, I have developed three kinds of vocabulary sets inspired by Smalley's work. They are:

- Cohesiveness, to assess the relationship between timbral emotional valence and the emotions conveyed by the lyrics;
- Attachment, to consider the relevance of the expressed emotions to the listener's experience; and
- Emotional Map, which concerns the emotional themes conveyed within a song.

7.4.1 Cohesiveness.

Cohesiveness is a vocabulary set that allows the analyser to explore the synergies between an emotionally valenced vocal timbre and emotionally valenced lyrics, ranging from vocal timbre and lyrics being completely aligned, to being completely misaligned (see Figure 7.4). The state of synergy of vocal timbre and lyrics may sit anywhere along this scale. This vocabulary set also allows the analyser to extrapolate from this state of synergy and to make judgments about its potential impact on emotional perception. For example, if one was to assess the vocal timbre and lyrics of Country Joe and the Fish's “The "Fish" Cheer/I-feel-like-I'm-fixin'-to-die rag” (McDonald, 1967, side 2, track 1), one may identify the vocal timbre as being upbeat and happy, and the lyrics as being negative and sad. This relationship may be described as misaligned in terms of Cohesiveness. It is possible then to extrapolate from this to say that such conflicting emotions create an unsettling feeling. The emotional message is ambiguous and the listener must continually reassess the vocal timbre and lyrics to determine what message is being conveyed by the performer.

Cohesiveness also provides a tool for describing the relationship between emotive content in vocal timbre and words in a way that is consistent across analyses. Of course, while the state of cohesiveness may change between analyses (in one analysis the vocal timbre and lyrics may be misaligned, in another analysis they may be aligned), the *pool of descriptors* outlined in Cohesiveness does not change between analyses. In this way, Cohesiveness provides a standardized tool to be used across analyses to assess the synergies of vocal timbre and lyrics and make judgments about how these may impact emotional perception.

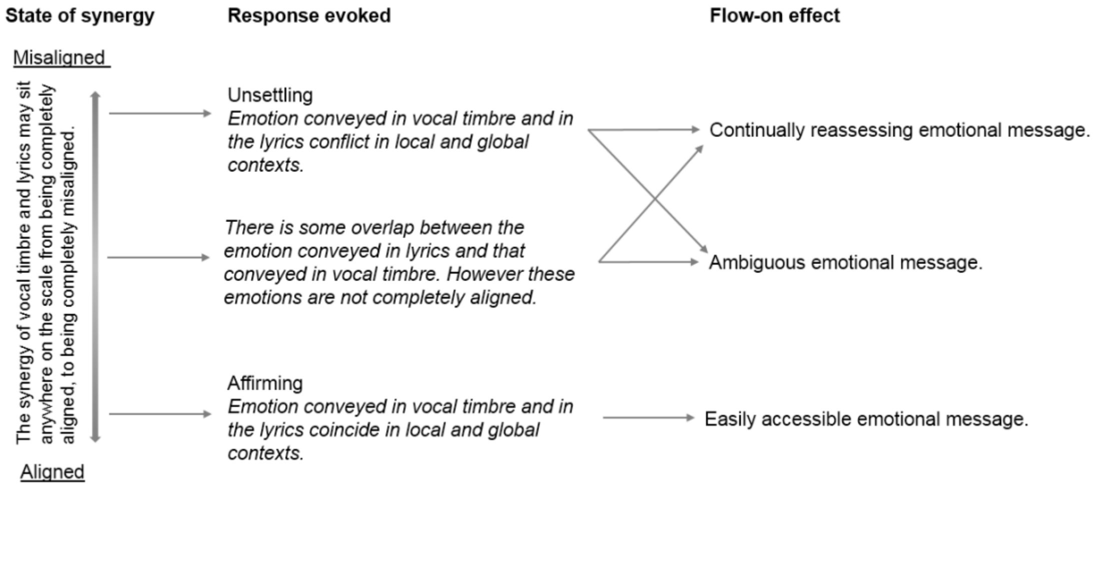

State of synergy
Response evoked
Flow-on effect
Misaligned
The synergy of vocal timbre and lyrics may sit anywhere on the scale from being completely aligned, to being completely misaligned.
Aligned
Unsettling
Emotion conveyed in vocal timbre and in the lyrics conflict in local and global contexts.
There is some overlap between the emotion conveyed in lyrics and that conveyed in vocal timbre. However these emotions are not completely aligned.
Affirming
Emotion conveyed in vocal timbre and in the lyrics coincide in local and global contexts.
Continually reassessing emotional message.
Ambiguous emotional message.
Easily accessible emotional message.

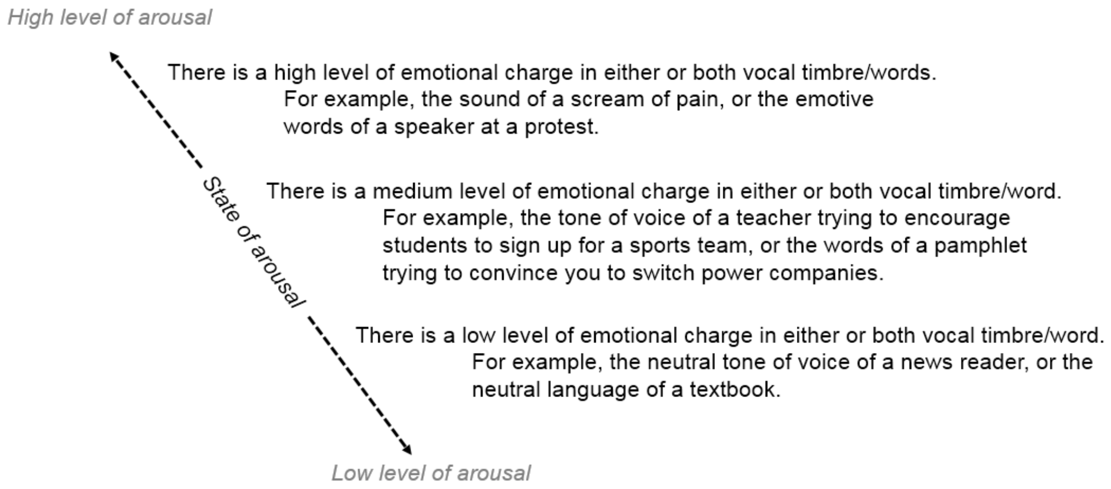
Figure 7.4. Cohesiveness. Vocabulary set to describe the relationships between the emotion in timbre and that in the lyrics, and the flow-on effects of such relationships.

7.4.2 Attachment.

Figure 7.5 depicts the vocabulary set of Attachment. Attachment provides a tool to assess how aroused a listener is likely to be by the emotion expressed by a performer through vocal timbre and/or lyrics, and to extrapolate from this to assess how closely connected the listener is likely to feel to the expressed emotion. Attachment is based on the idea that listeners are likely to feel a stronger response to vocal timbre/lyrics when these stimuli are emotionally valenced, immediately connected with the human experience, and likely to have been experienced by listeners.

At one end of this vocabulary set the listener may likely experience a high level of arousal (top left corner in Figure 7.5). If so, the listener may feel closely connected with the emotion a performer expresses through vocal timbre/lyrics. For example, a scream is an emotionally valenced sound which could signify danger and therefore arouse fear in the listener. Highly arousing emotions such as this are likely to have a strong impact on emotional perception. As one moves down the vocabulary set, the state of emotional arousal decreases such that at the other end of the vocabulary set (bottom right corner) the listener may likely experience a low level of emotional arousal. Here the vocal timbre/lyrics are mostly neutral in nature. For example, the neutral tones of a train announcer. The state of arousal evoked by vocal timbre/lyrics may sit anywhere along the two points of high and low arousal.

Attachment is like Cohesiveness in that, while the specific state of attachment may change between analyses (the emotional message in this analysis may be highly arousing, the emotional message in that analysis may evoke little arousal), the pool of descriptors outlined here do not change.



There is a high level of emotional charge in either or both vocal timbre/words. For example, the sound of a scream of pain, or the emotive words of a speaker at a protest.

There is a medium level of emotional charge in either or both vocal timbre/word. For example, the tone of voice of a teacher trying to encourage students to sign up for a sports team, or the words of a pamphlet trying to convince you to switch power companies.

There is a low level of emotional charge in either or both vocal timbre/word. For example, the neutral tone of voice of a news reader, or the neutral language of a textbook.

Figure 7.5. Attachment. Vocabulary set to assess the likely level of emotional arousal experienced by the listener.

7.4.3 Emotional map.

The Emotional Map is not so much a tool for conveying vocal timbre exclusively, rather it is designed to assist the analyser to “keep track” of the emotional themes as they are identified. Unlike the previous two vocabulary sets, the Emotional Map is not predefined but provides a template which can be “filled in” as the analysis progresses. In this way, the aim of the Emotional Map is to add clarity to analysis by allowing for the cataloguing of emotional themes present within vocal timbre and lyrics.

In the Emotional Map, the primary emotional themes the analyser believes to be present within the song can be documented in the section labelled “primary emotional theme” (Figure 7.6). If the analyser identifies more than one primary emotional theme within a song, then several Emotional Maps may be used in a single analysis to indicate this. This may be particularly helpful for when conflicting, or similar but distinct, emotional messages are present within a song (e.g., happy and sad, sad and fearful).

Possible ways to identify the emotional themes

The assessment of both vocal timbre and lyrics throughout the analysis will assist the analyser in identifying the primary emotional themes present within a song. In regard to vocal timbre, one should assess the acoustic features of a vocal timbre, and the context in which these features occur. For example, through the application of the Vocal Timbre Features discussed above, the analyser may identify a vocal timbre as being very breathy. This may indicate sadness, or may create a sense of intimacy. To make a judgment about which, the analyser would then need to assess the context. For example, how vocal timbre is used throughout a song (is it only breathy in the verse, but strong and resonant in the chorus?), and how vocal timbre is used in relation to other musical features (do the lyrics provide clues?). In this way, an informed judgment which can be musically justified may be made. In regard to lyrics, the analyser should assess the potentially emotional message conveyed by the words. In some cases (such as in the case studies presented in Chapter 8), these may be quite straight forward. In other cases, the lyrical message may be more ambiguous and the analyser may need to engage in an in-depth analysis of the lyrics to determine emotional themes, or document different emotional themes through several sets of emotional maps.

Throughout the analysis, an analyser may identify certain nuances of a primary emotional theme. These more detailed descriptors can be documented in the areas labelled *a*, *b*, *c*, *d*, *e*, *f*, etc. in Figure 7.6. For example, if *sadness* is identified as a primary emotional theme, then some descriptors may be *grief*, *anguish*, *pain*, and so on. In this way, the Emotional Map documents both primary emotional themes, and more detailed descriptors thus allowing one to "map" the emotionality of a song as the analysis is conducted.

A short note is necessary here on the relationship between my use of the term *primary emotion* and any other theory of basic human emotions. As discussed in section 1.3, this thesis has not restricted itself to any predetermined set or model of emotions (see section 1.3 for reasons why). Therefore, the Emotional Map too is not based on a prescribed set of basic human emotions as this may risk being too restrictive and may actually be fundamentally inappropriate. It is not the object of this thesis (nor is it necessary) to make theoretical judgments about the link between basic emotions and musical ones. I broach this topic only to justify my decision not to base the Emotional Map on any such predetermined set: theories about these emotions themselves continue to develop, and the relationship of these emotions to music is not always clear.

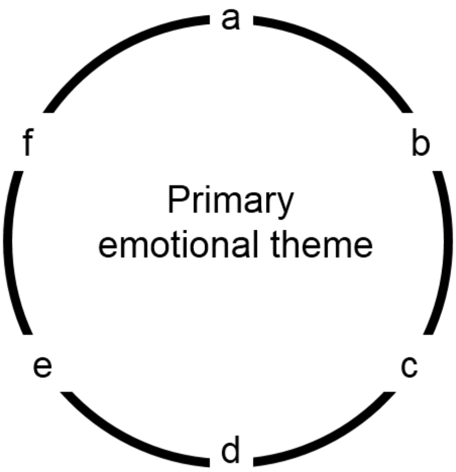


Figure 7.6. Emotional Map. Vocabulary set to lay out the primary emotional themes of a song (inserted into the section labelled *primary emotional theme*) and to describe the nuances of themes (listed on the places labelled *a, b, c,* and so on).

7.5 Conclusion

In this chapter, I have outlined a new analytical technique for vocal timbre in popular vocal song. This technique draws on several processes. These are:

- Selecting the recording to be analysed.
- Identifying and describing the acoustic features of a vocal timbre, and using these to determine emotional valence. This is achieved by making use of the Vocal Timbre Features and spectrographs to examine three main components of a vocal timbre: onset, sustain, and termination.
- Combining the aural experience (i.e., recording) and visual tools (Vocal Timbre Features, spectrographs, lyrics, and audible breaths) to produce an audiovisual description of the vocal timbre being analysed.
- Assessing the relationship between the emotions evoked within a listener by emotional cues within a timbre and those conveyed by the lyrics, the relationship between this emotion to a listener's everyday experience, and the likely consequences of this relationship on a listener's emotional perception of the vocal line as a whole.

This analytical technique yields several original contributions to the field of vocal timbre studies. First, it offers a method for presenting audiovisual descriptions of vocal timbre (as discussed in section 7.3.3). This is achieved by combining the aural experience (played back from the recording) with several descriptive tools into one cohesive presentation.

Second, the development of the Vocal Timbre Features provides a system of symbols for annotating a vocal timbre in an audiovisual way, and for succinctly describing vocal timbre on the page. In other words, the Vocal Timbre Features provide a set of graphic markers which can be used to represent the aural features of a vocal timbre.

Third, this approach allows one to explore the relationship and interactions between the emotion conveyed through vocal timbre and that conveyed through lyrics—a relationship that is supported by the results of the reception tests described in Part II. It also allows the analyser to then evaluate how this relationship may create emotional meaning. Evidence to support the underlying hypothesis that vocal timbre can impact a listener's perception of emotion in lyrics can be found in Part II.

In the next chapter, the application of this technique will be demonstrated through two case studies. Areas in which this technique may be developed in future research are outlined in Chapter 9.

8 Case Studies

8.1 Introduction

Having outlined the methodology of the approach proposed in this thesis, namely analysing how emotion conveyed in vocal timbre impacts emotional perception of sung words (see Chapter 7), this chapter will now present two detailed case studies demonstrating how the analytical technique could be applied.

The repertoire choices for the case studies are selected from the broad category of popular vocal songs which this technique has been primarily designed to analyse. These songs are particularly suitable for the kind of analysis proposed here as they tend to rely heavily on the vocal line to convey meaning and emotion, and the lyrics are usually easily intelligible. The specific songs chosen as case studies in this chapter have quite clear and unambiguous lyrical messages. They both speak of heartbreak, love, and loss—universal themes. I have chosen these songs precisely for this reason. This is because this chapter aims to demonstrate how this analytical technique may be applied, something that is easier to show with case studies where the lyrical message is relatively clear and accessible. In the future, this technique may be developed further to include different categories of popular music, and to explore more lyrically ambiguous songs. However for the time being examining songs that foreground the vocal line and that convey an accessible lyrical message seems to be a logical starting point for examining the role of vocal timbre in the creation of emotional meaning. Another reason these songs are appropriate is because they tend to place the vocals at or near the centre of the mix (see Chapter 7, discussion on analytical scope) which serves to emphasise the vocal line.

The terminology, tools, and techniques used in the case studies in sections 8.2 and 8.3 are explained in depth in Chapter 7. Each case study is divided into two parts. The Preparation section presents information about the song being analysed, and sets out other material necessary for the analysis. The second section offers the analysis proper.

8.2 Gotye's "Somebody That I Used to Know"

The first case study to be addressed is Gotye's "Somebody That I Used to Know" (De Backer, 2011, track 3), featuring Kimbra. Written in 2011, this song won several awards (including the Triple J Hottest 100, and Aria Award for Song of the Year and Best Video). The song has sold over 770,000 copies in Australia, 1,500,000 copies in the United Kingdom, and 7,900,000 copies in America, arguably making it Gotye's best-known work to date. These figures suggest that "Somebody That I Used to Know"—hereafter, "Somebody" (De Backer, 2011)—speaks in some way to a large number of listeners. Certainly, its themes of love and heartbreak have an element of universality to them, making it a good candidate for the case study presented here.

8.2.1 Preparation.

The recording

The recording to be used as the basis of analysis in this case study is:

De Backer, W. (2011, track 3). Somebody that I used to Know [Recorded by Gotye, featuring Kimbra]. On *Making mirrors* [CD single, digital download, 7" vinyl (promotional only)]. Merricks, Australia: Eleven.

Descriptive tools

Audiovisual Attachment 1 (AV1 for short) provides an audiovisual description of vocal timbre in “Somebody” (De Backer, 2011, track 3). To produce these video descriptions, I have used the following software: Audacity® (Audacity Team, 1999–2015) to produce the scrolling spectrographs, and ScreenFlow (Telestream, 2016) to combine these spectrographs along with other visuals.[18] AV1 combines several descriptive tools acting simultaneously in one presentation. First, the audio (from the isolated vocal track) is synced with the spectrograph. Then lyrics are added so

[18] Built-in ScreenFlow text was used to produce the written lyrics, and build-in ScreenFlow shapes/colours to annotate the Vocal Timbre Features/Audible Breaths.

that one can more easily follow the song. Next, places where audible breaths can be heard are noted. Finally, the Vocal Timbre Features are added and annotated, giving an audiovisual description of the vocal timbre. The Vocal Timbre Features are listed in Figure 7.2. The result of combining these tools is an annotated, audiovisual description of Gotye and Kimbra's vocal timbres.

While completing this kind of video description is, at present, very time consuming (it is hoped that part of this process can be automatized in the future), it does have the advantage of allowing one to become very familiar with the different features of the vocal timbre and lyrics being analysed. It also allows the analyser to hear the lyrics in the context of the vocal timbre, and to hear how vocal timbre/lyrics in one section relates to those in other sections (e.g., how does vocal timbre change from one chorus to the next?). Therefore, while the resulting video description is useful in analysis, the process of annotation is just as useful as it can assist the analyzer in "getting to know" a vocal timbre in minute detail.

Analytical structure

I have chosen to structure the analysis according to key sections within the song. The role of the voice is the main factor in delimiting these sections. They are*: Instrumental Only, Gotye Solo, Kimbra Solo, and Vocal Duet*. Figure 8.1 shows a timeline of the sections.

The Instrumental Only section is characterised by the absence of the vocal line. Layering occurs in the instruments, with one solo instrument generally being present in the centre of the track. Vocal timbre and lyrics can, of course, not be analysed in the Instrumental Only sections as these sections do not include a vocal line.

The Gotye Solo and Kimbra Solo sections are characterised by the presence of a solo vocal line over several layers of instrumental accompaniment. In these sections, the musical features have been arranged such that the vocal line is usually centred in the track. For example, in the opening verse, Gotye's voice can be heard in front of the guitar, and the drums can be heard off to the right. This arrangement places emphasis on the vocal line, foregrounding vocal timbre and lyrics in these sections.

Also in these sections, the vocal line is usually louder than other musical features. Further, the vocal line tends to contain variation, which has the effect of capturing attention. Certainly, there are short instrumental interludes that add some variation to the accompaniment (e.g., 0'33"–0'34", 0'36"–0'38", 0'40"–0'41", 0'46"–1'03", 1'09"–1'11", 1'15"–1'18", 1'32"–1'33", 1'36"–1'37")[19], but these do not occur with the vocal line, instead they tend to function as punctuation between lines.[20]

[19] In the times presented in this analysis, the single quotation mark (') represents minute/s and the double quotation mark (") represents second/s.

[20] There is some instrumental variation which occur simultaneously with the vocal line in the chorus—for example, a twangy guitar motive can be heard intermittently in the first chorus (beginning 1'33''). However, these motives are mediated by the arrangement and layering (e.g., they are placed behind the voice in the mix), meaning that, while they play off the voice, they do not compete with it. Therefore, in general, when the vocal line is present in this section, it contains variation which sets it apart from other instrumentation and draws the listener's attention.

The Vocal Duet section is similar to the Gotye/Kimbra Solo sections, however rather than being characterised by a single voice, these sections tend to feature the layering of several vocal tracks at once. While the vocal line is still foregrounded, I have made the decision to distinguish between the duet and solo sections for two reasons. First, Kimbra's vocals in this section consist mainly of nonwords, placing more emphasis on her sound and less on her lyrical (linguistic) message. While the emotional meaning in vocal timbre can certainly be examined at this point, there is less to discuss in relation to emotional words. Second, the layering of voices in the duet section creates dense, polyphonic vocal harmonies, resulting in the vocal passages in the duet being denser than in any previous section. This has the effect of foregrounding the sound of the voice as it creates a sudden, unexpected, increase of sound which is aurally quite striking, especially since this is also the first time such vocal layering has been heard in "Somebody" (De Backer, 2011, track 3). Overall, in the duet section the analysis of vocal timbre would be appropriate and meaningful, however, due to the inclusion of nonwords and the density of this passage, care would need to be taken when interpreting the synergies between the emotions conveyed through vocal timbre and that expressed in lyrics.

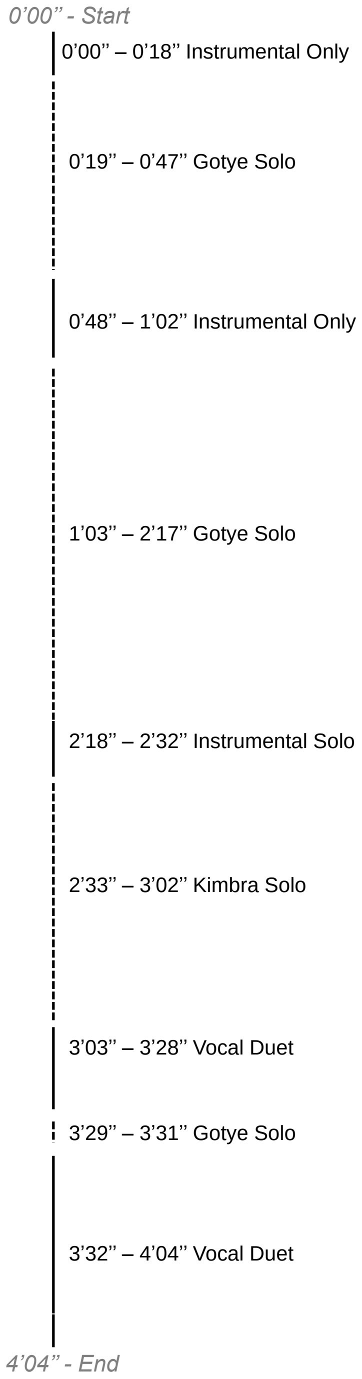

0’00” - Start
0’00” – 0’18’’ Instrumental Only
0’19” – 0’47’’ Gotye Solo
0’48” – 1’02’’ Instrumental Only
1’03” – 2’17’’ Gotye Solo
2’18” – 2’32’’ Instrumental Solo
2’33” – 3’02’’ Kimbra Solo
3’03” – 3’28’’ Vocal Duet
3’29” – 3’31’’ Gotye Solo
3’32” – 4’04’’ Vocal Duet
4’04” - End

Figure 8.1. Timeline of “Somebody” (De Backer, 2011, track 3) showing the different sections of the song (Instrumental Only, Gotye Solo, Kimbra Solo, Vocal Duet) and when they occur. The single quotation mark (‘) represents minute/s and the double quotation mark (“) represents second/s.

8.2.2 Analysis.

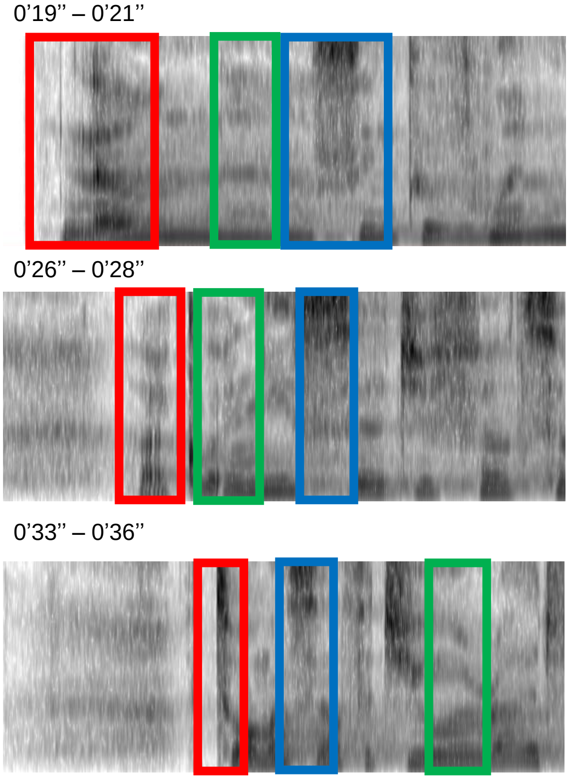
This section will draw on the tools introduced above and in Chapter 7 to conduct an analysis of vocal timbre in “Somebody” (De Backer, 2011, track 3). The analysis will be structured according to the order in which different sections appear in the song (see Figure 8.1). The times given for specific events are in reference to Audiovisual Attachment AV1. One should read this analysis in conjunction with this attachment.

Certain emotional themes and nuances will be identified throughout this analysis. As they are identified they will be incorporated into the Emotional Map (Figure 8.16). In this way, the emotional themes/nuances which are identified within “Somebody” will be presented in a visually succinct way at the end of this analysis.

8.2.2.1 Gotye solo: First two verses (0’19’’–0’47’’ and 1’03’’–1’32’’) and first chorus (1’34’’–2’17’’).

First two verses—0’19”–0’47” and 1’03”–1’32”

Gotye’s is the first voice heard in “Somebody”, opening the song with a soft, plain, and generally smooth vocal timbre. Gotye oscillates between his low and medium registers, which allows him to maintain a lax tone. In general, these vocal timbre features indicate a neutral emotive state—Gotye’s voice is neutral and relaxed, the voice of someone who is in control. However, some vocal nuances betray the underlying emotion. One such nuance is onset. Gotye's first vocal utterance (0'19") begins with a creak onset. This trend continues throughout the first two verses, with creak onsets tending to occur in the first few words of the phrases. Spectrographs of some of these onsets, shown in Figure 8.2, provide visual confirmation of this. Here, one can see the use of creak onsets at the start of the phrase, with other onsets (e.g., aspirate and simultaneous) being used thereafter. These creak onsets are noisy and rough, and this is indicated in the spectrographs by the presence of dense harmonics occurring throughout the spectrum. Because of the potential connection between noise and negative emotions (see Chapter 6, Discussion, and Chapter 7, section 7.3.1), this noise may suggest a negative emotional state, perhaps evoked by the reliving of painful memories. In fact, sadness has been identified as a primary emotional theme in this song, giving rise to a number of related emotional nuances, this is mapped in Figure 8.16. The lyrics support the idea that the character Gotye is portraying is reliving such a memory as they describe a sad event (indicated by words in the lyrics such as "lonely", and "ache") that occurred in the past (indicated using past tense such as "were" and "was").



0’19’’ – 0’21’’

0’26’’ – 0’28’’

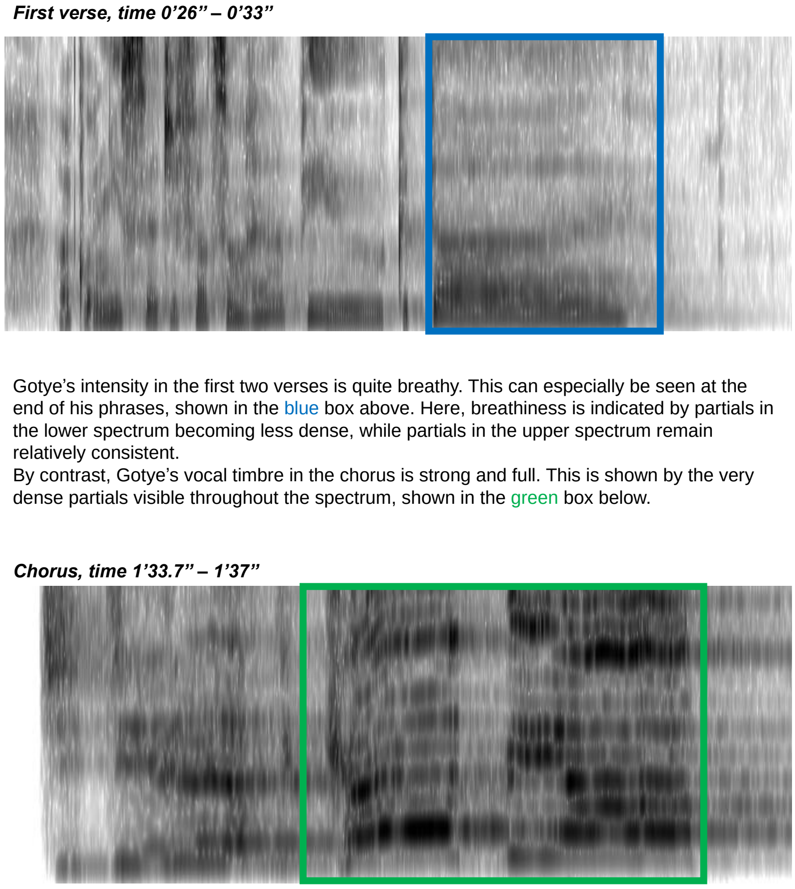
0’33’’ – 0’36’’

Creak onsets, shown in the red boxes, are indicated by the presence of darker partials occurring throughout the spectrum.

They are distinct form aspirate onsets, shown in the blue boxes, which are indicated by darker partials in the upper spectrum, and from simultaneous onsets (shown in the green boxes) which are indicated by darker partials in the lower part of the spectrum and lighter partials in the upper spectrum.

Figure 8.2. Spectrographs of some of Gotye’s creak, aspirate, and simultaneous onsets in the first verse of “Somebody”. The single quotation mark (‘) represents minute/s and the double quotation mark (“) represents second/s.

The audible inclusion of breath is another cue which suggests that Gotye's character is not as composed as his words might suggest (i.e., it gives the impression that Gotye's character is likely experiencing a strong emotional response to something). Breath can be heard occurring in two ways. First, breath can be heard in the vocal timbre's sustain, particularly towards the end of phrases. This is a characteristic of Gotye's vocal timbre in the first and second verses, and it tends not to be seen later in "Somebody". Figure 8.3 shows an example of this breathy sustain. Here, a phrase from the first verse (0'26"–0'33") is contrasted with one from the chorus (1'33.7" – 1'37"). In the first verse phrase, the breathiness can be seen by the decrease in harmonics in the lower spectrum, while the harmonics in the upper spectrum remain relatively consistent. In the chorus, however, the harmonics remain dense and steady throughout the spectrum. The breathy voice heard in the first verse is controlled but lacking energy. Such qualities may convey a sense of emptiness (perhaps generated by the underlying sadness present in the first verse, this relationship is mapped in Figure 8.16), an emotion which is low in energy (compared to fear, for example, which is high in energy) and, potentially, negative (suggested by the noisy quality of the breath).



First verse, time 0’26” – 0’33”

Gotye’s intensity in the first two verses is quite breathy. This can especially be seen at the end of his phrases, shown in the blue box above. Here, breathiness is indicated by partials in the lower spectrum becoming less dense, while partials in the upper spectrum remain relatively consistent.

By contrast, Gotye’s vocal timbre in the chorus is strong and full. This is shown by the very dense partials visible throughout the spectrum, shown in the green box below.

Chorus, time 1’33.7” – 1’37”

Figure 8.3. Spectrographs of the breathy sustain used by Gotye in the first verse, as compared to the stronger sustain of the chorus. The single quotation mark (‘) represents minute/s and the double quotation mark (“) represents second/s.

The second way in which air can be heard in the voice is through *audible intakes* of breath. These occur often throughout the first two verses (see AV1 at 0’26’’, 0’33’’, 0’37’’, 0’41’’, 1’03’’, 1’10’’, 1’18’’, 1’22’’, and 1’26’’). Certainly, breathing is a natural part of singing, so these audible breaths may be considered by-products of vocal production. However, this study is more concerned with listener perception than composer/performer intent. Therefore, the fact that these breaths can be heard, and that they have the potential to impact listener perception, is sufficient for them to be included in this analysis. Furthermore, if they were simply an unavoidable by-product captured in the recording process, these breaths could have been minimised or eliminated from the track during production. Therefore, it may be that they were retained to act as emotional indicators, or perhaps to retain the *natural* sound of the voice (we hear breath in the spoken voice, not hearing it in the sung voice may alienate the listener).

The frequency of these audible breaths may suggest to the listener that the character Gotye is inhabiting is suppressing some underlying emotions. For example, it is likely not necessary for a breath between the lines “So when we found that we could not make sense, Well you said that we would still be friends” (see AV1, beginning 1’18’’). Gotye does however breathe, as though his character is bracing himself against the emotions evoked by the memory he is reliving through telling his story. In this way, breath may evoke a sense of sadness (a primary emotional theme, see Figure 8.16) and pain (a nuance of this sadness, represented in Figure 8.16).

As can be seen from the above discussion, there are several different emotional cues present in Gotye's vocal timbre. The neutrality suggested by the lax, low, plain vocal features is contrasted with creak onsets and audible intakes of breath which evoke a sense of sadness and pain, and with the breathy sustain which evokes a sense of emptiness. The combination of such vocal timbre cues give the impression that Gotye's character is trying to distance himself from the underlying emotions he is feeling—emotions that are communicated to the listener via the lyrical content.

The lyrics of the first two verses are:

Verse 1

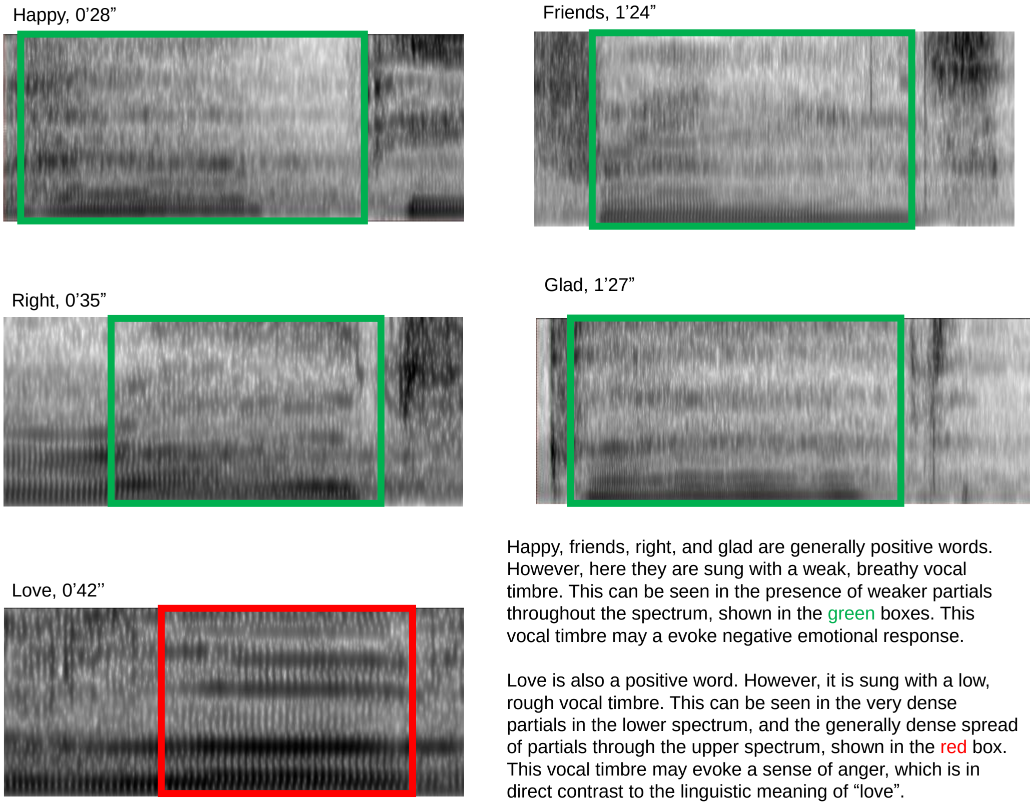
Now and then I think of when we were together.
Like when you said you felt so happy you could die.
Told myself that you were right for me,
But felt so lonely in your company.
But that was love and it's an ache I still remember. (De Backer, 2011, track 3)

Verse 2

You can get addicted to a certain kind of sadness.
Like resignation to the end, always the end.
So when we found that we could not make sense,
Well you said that we would still be friends.
But I'll admit that I was glad it was over. (De Backer, 2011, track 3)

A sense of loss is at the centre of this message (another emotional nuance reflected in the Emotional Map, Figure 8.16). The character can recall the ache of love, the addiction to the sadness, and the angry but subdued resignation to the pain that would follow the breakup. Emptiness is also a focal point as Gotye's character recalls the lies he told himself to prolong the relationship ("told myself that you were right for me"), even though he knew it would not last ("but felt so lonely in your company"). In the end, he reveals that the only joy he found in the relationship was in its ending ("I was glad it was over"). This lyrical message, which conveys an overall emotion of sadness, is aligned with that of the vocal timbre.

However, there are instances of vocal timbre/lyrics being misaligned in the first two verses. In these instances, words which would typically be considered as having a positive emotional meaning are sung with a negative vocal timbre: "happy" (0'28''), "right" (as in correct, "right for me") (0'35''), "love" (0'42''), "friends" (1'24''), "glad" (1'27'') (see AV1). Generally, these words are sung with a breathy and weak vocal timbre, which may evoke a sense of sadness and loss. The word "love", however, receives a different treatment. It is low and rough, almost a growl, which may evoke a sense of anger (another primary emotional theme identified in "Somebody", which is mapped in Figure 8.16). This contrast can be seen in the spectrographs presented in Figure 8.4. While the presentation of these positive words with a negative vocal timbre is arousing and unsettling—the emotionally valenced vocal timbres and words arouse our emotions, the fact that these arousing emotions are misaligned means that we must continue to reassess vocal timbre/lyrics to decipher the "true" emotional message—particularly in the treatment of the word love. This is because *love* is a word that is typically associated with intimacy and tenderness. The roughness with which this word is sung here betrays this usual emotional association, taking a positive and intimate word and infusing it with anger, pain and sadness. This treatment of “love” (see AV1, 0’42”) creates a sense of unpredictability (Gotye’s character was sad, now he is angry) which in turn generates a feeling of danger—it is difficult to know what form the character’s emotional state will take next (danger is mapped as a nuance of anger in Figure 8.16).



Happy, 0’28”

Friends, 1’24”

Right, 0’35”

Glad, 1’27”

Love, 0’42’’

Happy, friends, right, and glad are generally positive words. However, here they are sung with a weak, breathy vocal timbre. This can be seen in the presence of weaker partials throughout the spectrum, shown in the green boxes. This vocal timbre may a evoke negative emotional response.

Love is also a positive word. However, it is sung with a low, rough vocal timbre. This can be seen in the very dense partials in the lower spectrum, and the generally dense spread of partials through the upper spectrum, shown in the red box. This vocal timbre may evoke a sense of anger, which is in direct contrast to the linguistic meaning of “love”.

Figure 8.4. Spectrographs illustrating the contrast between the rough vocal timbre used to sing the word “love” and the breathy vocal timbres used to sing other generally positive words (“happy”, “friends”, “glad”, and “right”). *The single quotation mark (‘) represents minute/s and the double quotation mark (“) represents second/s.*

The contrast of these brief instances of misalignment with the otherwise aligned vocal timbre/lyrics only serve to heighten the overall emotional message of pain, sadness, and emptiness expressed in the first two verses. In a general (i.e., global) sense, the vocal timbre and lyrics tend to be aligned, creating an emotional message that is mostly easily accessible for the listener (see Figure 8.5). The brief word-to-word (i.e., local) violations in alignment serve to further affirm this message—Gotye's character is trying to put up a brave front, but the memory of the relationship causes him pain, therefore it makes sense that this pain would intermittently present itself in these brief, local, contexts. In other words, the violations are *contextually appropriate*. Additionally, these emotional themes (relationship breakdowns, putting on a brave front) are common experiences. Because these emotional themes have the potential to be relevant to the listeners' experience, the emotional message is also likely to be highly arousing (see Figure 8.6).

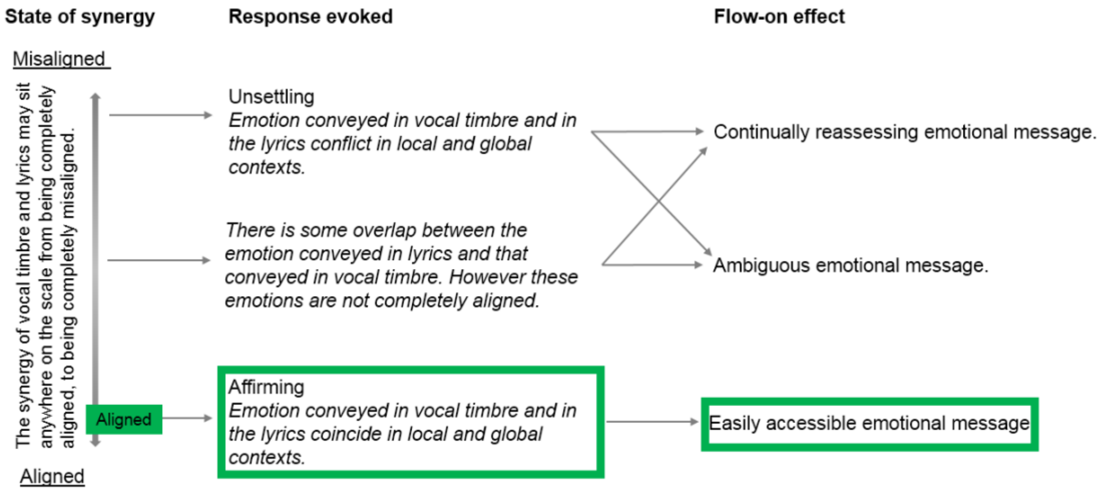

State of synergy
Response evoked
Flow-on effect
Misaligned
The synergy of vocal timbre and lyrics may sit anywhere on the scale from being completely aligned, to being completely misaligned.
Unsettling
Emotion conveyed in vocal timbre and in the lyrics conflict in local and global contexts.
Continually reassessing emotional message.
There is some overlap between the emotion conveyed in lyrics and that conveyed in vocal timbre. However these emotions are not completely aligned.
Ambiguous emotional message.
Aligned
Affirming
Emotion conveyed in vocal timbre and in the lyrics coincide in local and global contexts.
Easily accessible emotional message
Aligned

Figure 8.5. Cohesiveness in the first two verses of "Somebody" (De Backer, 2011, track 3). Highlighted sections show the synergies between vocal timbre and lyrics.

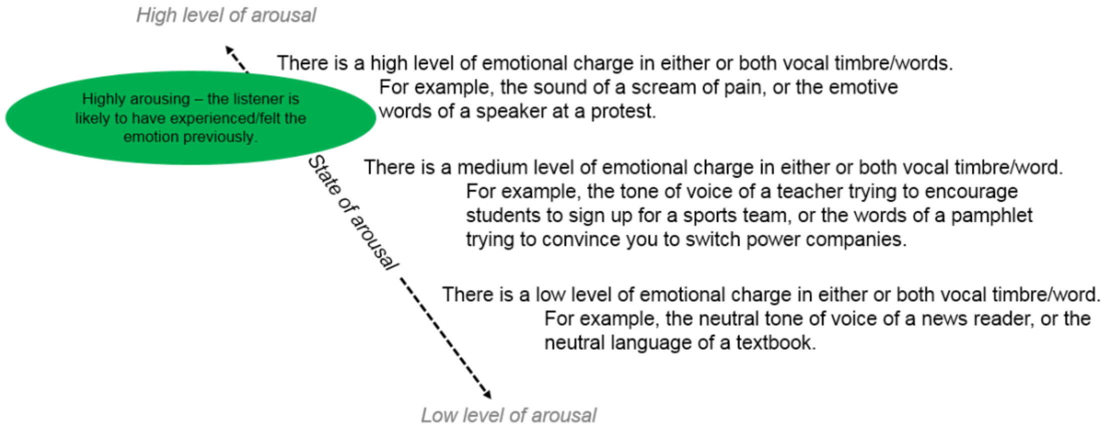
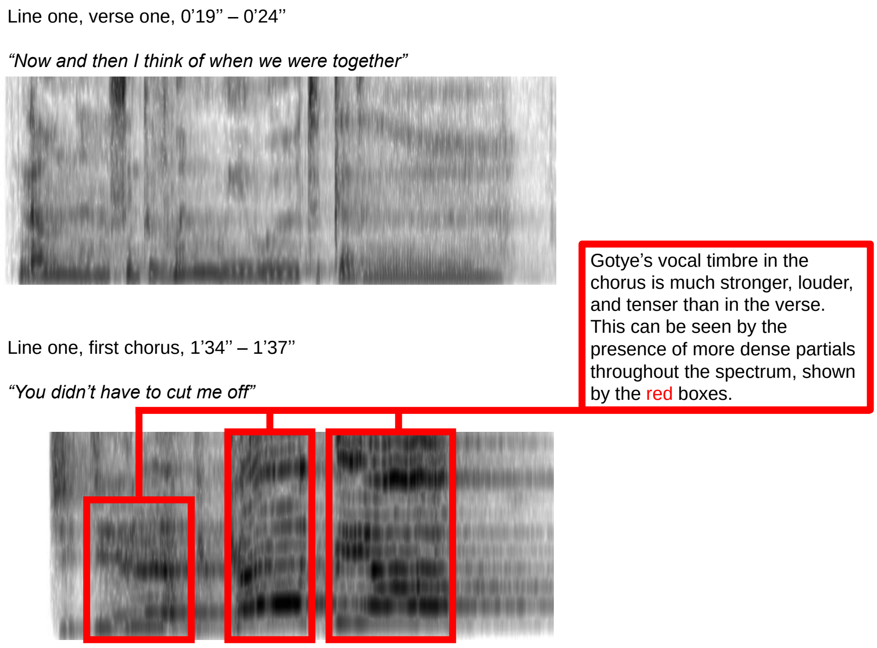

High level of arousal
There is a high level of emotional charge in either or both vocal timbre/words. For example, the sound of a scream of pain, or the emotive words of a speaker at a protest.
Highly arousing – the listener is likely to have experienced/felt the emotion previously.
State of arousal
There is a medium level of emotional charge in either or both vocal timbre/word. For example, the tone of voice of a teacher trying to encourage students to sign up for a sports team, or the words of a pamphlet trying to convince you to switch power companies.
There is a low level of emotional charge in either or both vocal timbre/word. For example, the neutral tone of voice of a news reader, or the neutral language of a textbook.
Low level of arousal

Figure 8.6. Attachment in the first two verses of "Somebody" (De Backer, 2011, track 3). The highlighted sections show the level of arousal evoked by the emotive content of vocal timbre and lyrics.

First chorus—1’34”–2’17”

There is a marked difference between the emotional message expressed in the first two verses, and that expressed in the first chorus (1’34”–2’17”). The breathy qualities present in the first two verses disappear, and instead Gotye’s vocal timbre becomes suddenly high, tense, and loud. This contrast can be seen in AV1 (verse beginning 0’19” compared to chorus beginning 1’34”) and in Figure 8.7. The volume of Gotye’s vocal timbre makes it sound as though he is shouting the lyrics, perhaps in *angry frustration* (anger can be frustration inducing, which is why it has been included in the Emotional Map, see Figure 8.16) of the injustice of Kimbra’s character, who we will soon meet, cutting him off. The register and tenseness also point to such emotions since rising tension and register are vocal qualities that can signify the presence of anger (as explored in Chapter 7, “[t]he sound that results from tension not only *is* tense, it also *means* ‘tense’ – and *makes* tense” (van Leeuwen, 1999, p.131))[21]—anger again possibly evoked by frustration at Kimbra’s character’s injustice.

[21] Anger is one emotion which may result in one feeling, and thus sounding, tense.



Line one, verse one, 0’19’’ – 0’24’’

“Now and then I think of when we were together”

Line one, first chorus, 1’34’’ – 1’37’’

“You didn’t have to cut me off”

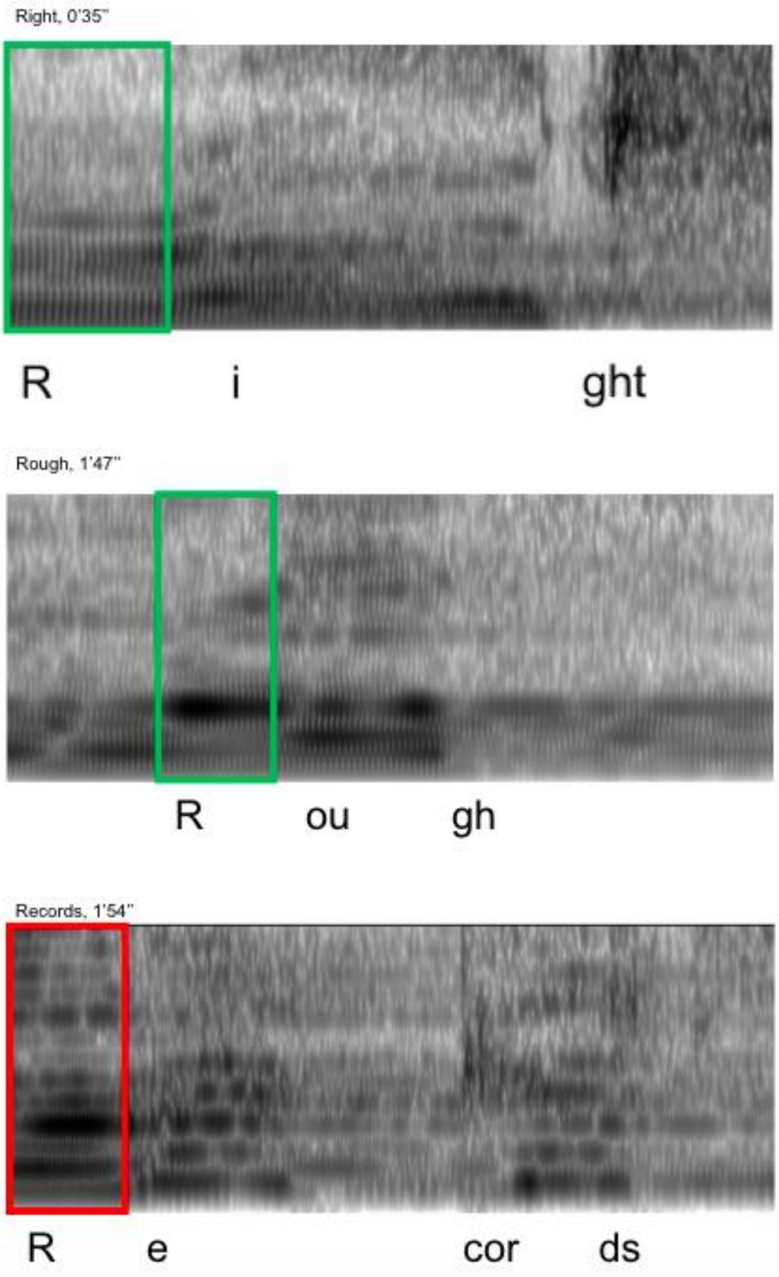
Gotye’s vocal timbre in the chorus is much stronger, louder, and tenser than in the verse. This can be seen by the presence of more dense partials throughout the spectrum, shown by the red boxes.

Figure 8.7. Spectrographs showing the contrast between Gotye’s softer, breathier vocal timbre in the verse compared to his stronger, louder, tenser vocal timbre in the chorus. The single quotation mark (‘) represents minute/s and the double quotation mark (“) represents second/s

Onsets also change from the verse to the chorus, with simultaneous onsets characterising the chorus. Because the vocal timbre is in general much higher, tenser, and louder, these simultaneous onsets sound very strong, and have an almost forced quality to them. This is again reminiscent of the feeling of shouting in angry frustration evoked by the sustain discussed above—shouted words are usually well enunciated and projected, as they are here in Gotye's vocal timbre. While simultaneous onsets are prevalent in the chorus, aspirate onsets can also be heard at various points. Some aspirate onsets in this section occur because of the natural pronunciation of the word. For example, in spoken English, certain word-initial sounds are produced by allowing breath to move through open vocal folds (e.g., voiceless sounds as described by Collins & Mees, 2013, p. 32). Therefore, it is likely that these same sounds will be aspirated (according to the definition provided in Chapter 7) in singing too. For example, aspiration and breath occur on the initial *ch* sound of "you" at 1'34" being pronounced in an informal, colloquial accent such that the word is said "*ch*-ou", the *s* of "*s*tranger" at 1'45", and the *s* of "*s*toop" at 1'49" (see AV1). However, aspirate onsets also occur on words which would not normally require one. For example, the word "records" (1'54") (see AV1) has an aspirate onset even though it would not normally require one in speech (see discussion of *r* sounds in glossary of Collins & Mees, 2013, p. 295–308).

And indeed, Gotye has shown in other sections that he can and does sing this "r" onset without aspiration. For example, "right" at 0'35", and "rough" at 1'47" (see AV1 and Figure 8.8). In both cases, Gotye employs simultaneous onsets, similar to how one would say these words in everyday speech. Therefore, that this aspirate onset occurs within a vocal timbre which otherwise mostly consists of simultaneous onsets, and that it occurs on a word which it does not necessarily have to, causes the lyric "records" to be emphasised in the aural experience (see AV1). The presence of this aspirate onset may indicate a brief, but salient, change in emotional message. The high, weak, and breathy quality of the vocal timbre that accompanies "records" suggests that Gotye's character has reverted to his earlier emotional state of sadness (see AV1). However, the speed with which this change occurs (the aspiration only lasts a second before a return to simultaneous onsets) also suggests volatility (perhaps evoked by the undercurrent of anger, another relationship documented in Figure 8.16). In other words, the sudden change in vocal performance suggests that Gotye's character is not in control of his emotional state.

Further, *record* is a noun that is not generally imbued with any emotional connotations. Therefore, it is less as though Gotye's aspiration here is attributed to the emotionality of the particular word, and more as though this aspiration is indicative of an uncontrollable, emotional outburst. This gives the effect that Gotye's character is, again, only just in control of his emotions (as was the case in verses one and two), creating an unsettling effect and causing the listener to be continually on the lookout for other emotional clues. Overall, Gotye's vocal timbre in the chorus is angry, but pained.



Simultaneous onsets are used on the words "right" and "rough". This is shown by the presence of darker partials in the lower spectrum, and lighter partials in the upper spectrum (green box). Such a spread of partials indicates the presence of a relatively strong fundamental note.

"Records", on the other hand, occurs with a more aspirated onset. This is shown by the darker spread of partials throughout the spectrum (red box) indicating a breathier, noisier, start to the note.

Figure 8.8. Spectrographs of three instances of "r" onsets in "Somebody" (De Backer, 2011, track 3). "Right" and "rough" tend to be sung with simultaneous onsets, with "records" tending to be more aspirated. The single quotation mark (') represents minute/s and the double quotation mark (") represents second/s.

The cause for this anger and pain may be found in the lyrical message, in which Gotye's character expresses a feeling of betrayal (making it a nuance of both of the emotional themes of the song, see Figure 8.16):

> But you didn't have to cut me off.
> Make out like it never happened and though we were nothing.
> And I don't even need your love,
> But you treat me like a stranger and that feels so rough.
> No, you didn't have to stoop so low.
> Have your friends collect your records and then change your number.
> I guess that I don't need that though.
> Now you're just somebody that I used to know.
> Now you're just somebody that I used to know.
> Now you're just somebody that I used to know. (De Backer, 2011, track 3)

Not only was he betrayed in love (he was "cut off" and she stooped so "low"), but his memory was also betrayed. In making "out like it never happened" and acting as though they "were nothing", Kimbra's character has taken from Gotye's character not only her physical possessions (her "records"), but also the memory of the relationship. For Gotye's character, this seems to be the worst kind of betrayal as memory is the one artefact of the relationship over which both parties may claim ownership. Betrayal is both rooted in anger (Kimbra's character is acting in a way that is not necessary, she "didn't have to cut him off") and in sadness (Kimbra's character doesn't recognise the love they did share, "you treat me like a stranger and that feels so rough") (this relationship between betrayal, and anger and sadness is depicted in Figure 8.16).

Overall in the chorus, the use of strong, simultaneous onsets contrasted with the occasional aspirated ones suggest barely concealed anger which slips, without warning, into sadness. The lyrical message, which is much more focused on betrayal, is slightly misaligned with the vocal timbre message of anger and frustration. These concurrent, yet slightly misaligned, messages (frustration and anger, contrasted with betrayal) create a sense of volatile tension within the music—the emotional message is ambiguous, meaning the listener must attend closely to vocal timbre and lyrics in order to understand, and prepare for, Gotye's character's next emotional state (see Figure 8.9).

As in the verse, the emotional message in the chorus is highly arousing. Gotye's vocal timbre and lyrics are emotionally valenced, and his overall message is closely connected to the human experience (i.e., likely to have been experienced by most people) (see Figure 8.10).

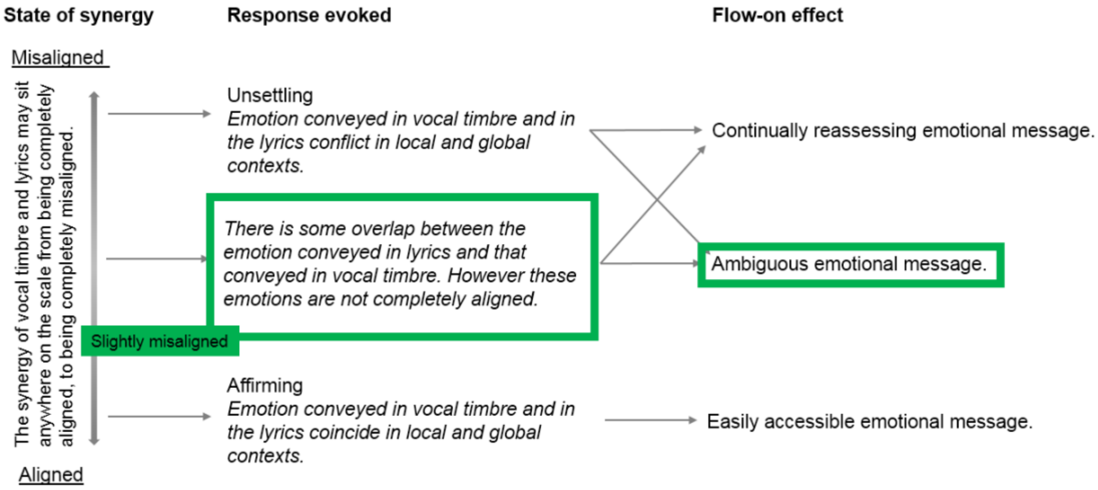

State of synergy
Response evoked
Flow-on effect
Misaligned
The synergy of vocal timbre and lyrics may sit anywhere on the scale from being completely aligned, to being completely misaligned.
Slightly misaligned
Aligned
Unsettling
Emotion conveyed in vocal timbre and in the lyrics conflict in local and global contexts.
There is some overlap between the emotion conveyed in lyrics and that conveyed in vocal timbre. However these emotions are not completely aligned.
Affirming
Emotion conveyed in vocal timbre and in the lyrics coincide in local and global contexts.
Continually reassessing emotional message.
Ambiguous emotional message.
Easily accessible emotional message.

Figure 8.9. Cohesiveness in the first chorus of "Somebody" (De Backer, 2011, track 3). Highlighted sections show the synergies between vocal timbre and lyrics.

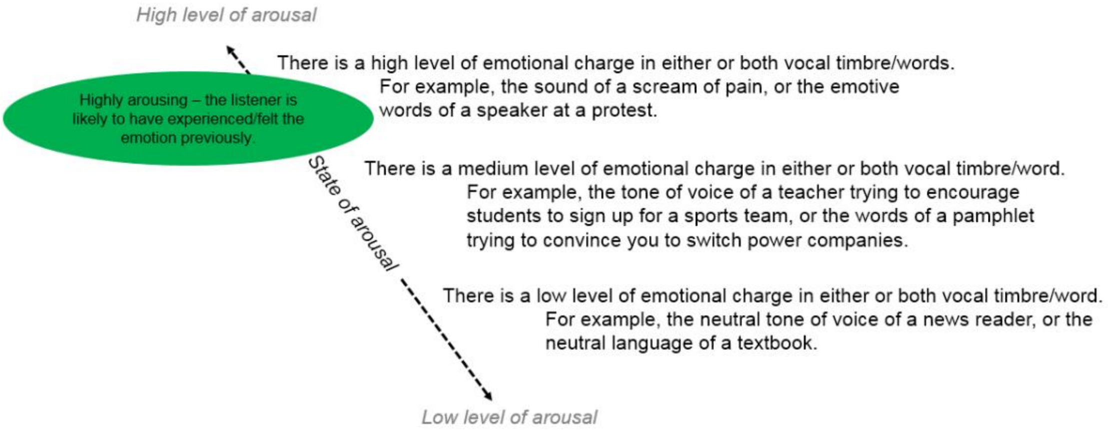

High level of arousal
Highly arousing – the listener is likely to have experienced/felt the emotion previously
There is a high level of emotional charge in either or both vocal timbre/words. For example, the sound of a scream of pain, or the emotive words of a speaker at a protest.
State of arousal
There is a medium level of emotional charge in either or both vocal timbre/word. For example, the tone of voice of a teacher trying to encourage students to sign up for a sports team, or the words of a pamphlet trying to convince you to switch power companies.
There is a low level of emotional charge in either or both vocal timbre/word. For example, the neutral tone of voice of a news reader, or the neutral language of a textbook.
Low level of arousal

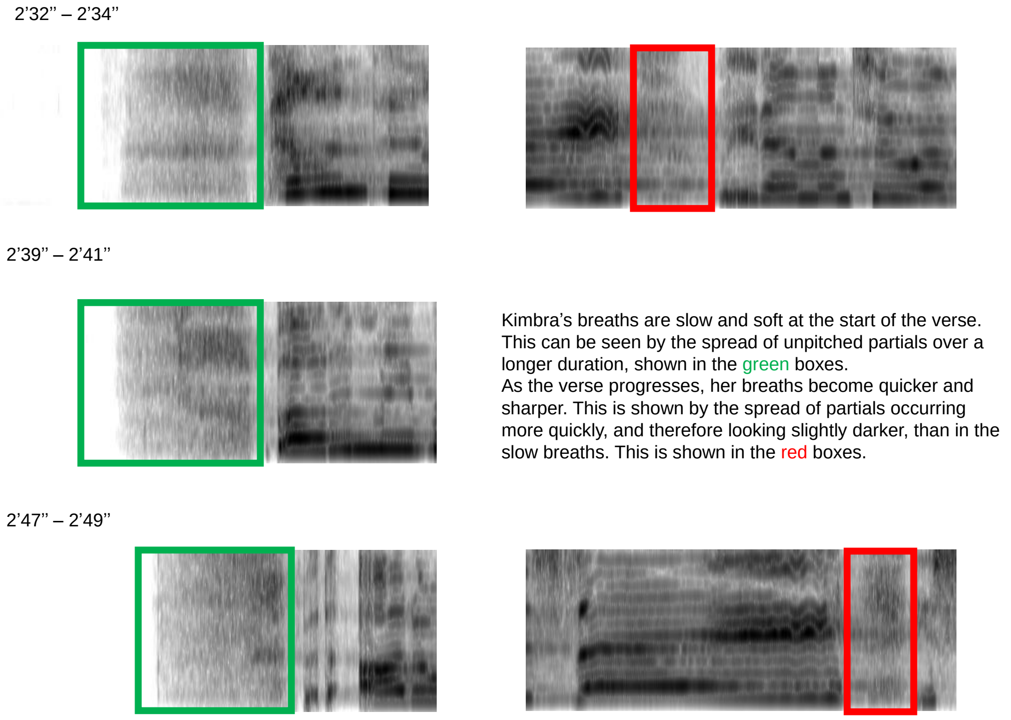
Figure 8.10. Attachment in the first chorus of "Somebody" (De Backer, 2011, track 3). The highlighted sections show the likely level of arousal evoked by the emotive content of vocal timbre and lyrics.

At the end of the first chorus, one is left feeling a sense of sadness and betrayal. The contrast between the weaker, more breathy opening, and the stronger, higher, and tenser chorus, together with the synergy of vocal timbre and lyric message moving from aligned to slightly misaligned, creates a slightly unsettling feeling (Gotye's character begins sad and disconnected (first two verses), but quickly becomes angry and frustrated (chorus)). At all stages of the song thus far, the message is easily relatable and taps the empathetic responses of the listener—most people have felt heartbreak/have been wronged in love (heartbreak and sadness often go hand in hand, as reflected in Figure 8.16). The oscillation of the emotional message of vocal timbre/lyrics between sadness (first two verses) and anger/frustration (chorus) in the global context (i.e., in the song thus far) creates a sense of ambiguity—what emotional state will Gotye's character inhabit next?

8.2.2.2 Kimbra only: Third verse—2'33''–3'02''.

Kimbra's vocal timbre bears some resemblance to Gotye's. The first similarity is the intake of breath. Kimbra's entrance at 2'33" is heralded, like Gotye's entrance at 0'19", by an audible breath. These breaths remain a feature throughout Kimbra's verse. As can be seen in AV1, Kimbra's initial breaths are soft and slow (see AV1, at 2'33", 2'40", 2'48"), however they soon become sharper and faster (see AV1, at 2'51", 2'55"). The spectrographs in Figure 8.11

illustrate this use of breath in more detail. Here, the sharper and faster intakes of breath are distinguishable by the spread of slightly darker, unpitched harmonics occurring relatively quickly. The slower intakes of breath at the start of the verse may suggest pain and vulnerability (emotions induced by her sadness, shown in Figure 8.16)—Kimbra’s character must draw breath to steady herself against the pain aroused by the lyrics, but she does this softly, unobtrusively, and timidly. As the verse progresses, however, Kimbra’s breaths become sharper and faster. The pain has not disappeared, but the vulnerability is giving way to frustration and resolve (an emotional nuance that may be drawn from anger, illustrated in Figure 8.16). It is as though Kimbra’s character is strengthened by her lyrical message.



Kimbra’s breaths are slow and soft at the start of the verse. This can be seen by the spread of unpitched partials over a longer duration, shown in the green boxes.
As the verse progresses, her breaths become quicker and sharper. This is shown by the spread of partials occurring more quickly, and therefore looking slightly darker, than in the slow breaths. This is shown in the red boxes.

Figure 8.11. Spectrographs illustrating how Kimbra’s intake of breath becomes faster and sharper as the verse progresses. The single quotation mark (‘) represents minute/s and the double quotation mark (“) represents second/s.

Kimbra’s onsets follow a similar pattern. At the start of the verse, Kimbra’s onsets tend to be aspirated, with a few instances of glottal onsets present (see AV1, 2’33”). As the verse progresses, Kimbra’s onsets become more simultaneous (beginning 2’48”) and by 2’58” the onsets are almost exclusively glottal (see AV1). It would seem that this contrast in onset from 2’33”–3’58” is deliberate as there are only a few instances in these lyrics where aspirated and breathy onsets would be needed in speech in general (one only needs to speak these lyrics out loud to identify places where the onset is produced by breath passing through relaxed/open vocal folds). These are indicated in italics here:

> Now and then I think of all the *t*imes [emphasis added] you *s*crewed [emphasis added] me over.
> But *h*ad [emphasis added] me believing it was always *s*omething [emphasis added] that I'd done. (De Backer, 2011, track 3)

However, Kimbra aspirates more onsets than just this (see AV1, 2’33”–3’58”), creating the impression that such timbral nuances are not only intentional but also, potentially, meant to act as emotional indicators. Kimbra’s sustain follows a similar pattern to that of onset. At 2’33”, Kimbra’s vocal timbre is typically soft and lax. At 2’48” however, the tension begins to increase and at 3’00” Kimbra’s vocal timbre is both very loud and very tense.

The vocal timbre in verse three suggests two different emotional states: vulnerability and pain (rooted in sadness, possibly for the love that had been, illustrated in Figure 8.16) developing into resolve and frustration (rooted in anger, perhaps evoked by the memory of being controlled by Gotye's character, mapped in Figure 8.16). Kimbra's vocal timbre at 2'33" is quite noisy, suggesting a negative emotion. This noise, however, is different to that in Gotye's verses. Rather than being rough and loud, it is breathy and soft. It is as though Kimbra's character is whispering to herself rather than speaking to Gotye's character or the listener, she does not necessarily want to be heard. This timid delivery reinforces the sense of vulnerability evoked by the intakes of breath discussed above. This emotional state is not permanent, though. At 2'48", the increasing volume and tenseness, and the increasing strength and clarity of the onsets, create a different effect. This time Kimbra's character wants to be heard, it is almost as though she has resolved to be heard—she is shouting her lyrics in a vocal timbre that matches Gotye's in the previous chorus.

At 2'48" there is a key turning point in Kimbra's vocals. It is here that her soft, breathy vocal timbre begins to give way to a stronger sound. It is here too that Kimbra's lyrical message begins to change. At 2'33", Kimbra begins with:

> Now and then I think of all the times you screwed me over.
> But had me believing it was always something that I'd done. (De Backer, 2011, track 3)

The lyrical message here is defensive—Kimbra's character has been mistreated, "screwed" over, and, perhaps, been emotionally controlled (as suggested by the line "but had me believing it was always something that I'd done"). However, from 2'48", her message changes:

> But I don't wanna live that way.
> Reading into every word you say.
> You said that you could let it go,
> And I wouldn't catch you hung up on somebody that you used to know.
> (De Backer, 2011, track 3)

These lyrics convey Kimbra's character's desire to change her circumstances, and remove herself from a negative situation. This change in lyrical message is supported by a change in the emotion conveyed through her vocal timbre. At 2'48" Kimbra's vocal timbre is strong, and so are her words—she is asserting herself, linguistically and vocally, in a way she was not permitted to/not able to previously. By 2'55", "You said that you could let it go, and I wouldn't catch you hung up on somebody that you used to know" (see AV1), Kimbra's vocal timbre is yet stronger and more intense. Resolve is now mixed with anger and frustration as Kimbra's character realises that she has been wronged and misused—she is angry that it happened and resolved to never let it happen again.

Overall, in verse three, the vocal timbre and lyrics are aligned in their emotional message. As the vocal timbre becomes more assertive, so too do the lyrics. This affirms Kimbra's character's narrative—she is very clear and specific about how she has been mistreated, and so we too feel sure that this must be true. There is no ambiguity here, the message is easily accessible (see Figure 8.12). It is likely that this message is highly arousing for a listener as, like in the Gotye only sections discussed above, most people will have experienced anger at having been wronged in a relationship, and the sense of clarity that such anger can bring (see Figure 8.13).

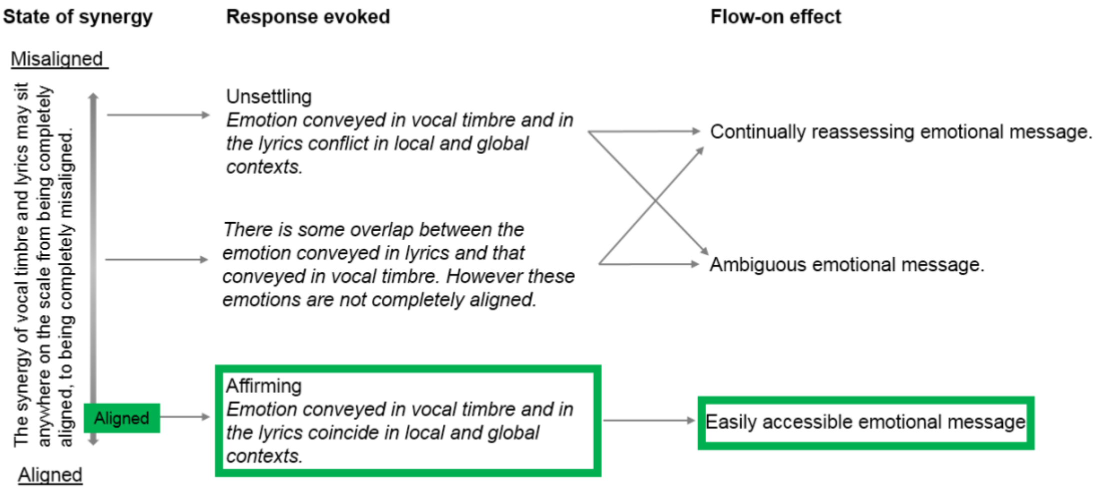

State of synergy
Response evoked
Flow-on effect
Misaligned
The synergy of vocal timbre and lyrics may sit anywhere on the scale from being completely aligned, to being completely misaligned.
Unsettling
Emotion conveyed in vocal timbre and in the lyrics conflict in local and global contexts.
Continually reassessing emotional message.
There is some overlap between the emotion conveyed in lyrics and that conveyed in vocal timbre. However these emotions are not completely aligned.
Ambiguous emotional message.
Aligned
Affirming
Emotion conveyed in vocal timbre and in the lyrics coincide in local and global contexts.
Easily accessible emotional message
Aligned

Figure 8.12. Cohesiveness in verse three of "Somebody" (De Backer, 2011, track 3). Highlighted sections show the synergies between vocal timbre and lyrics.

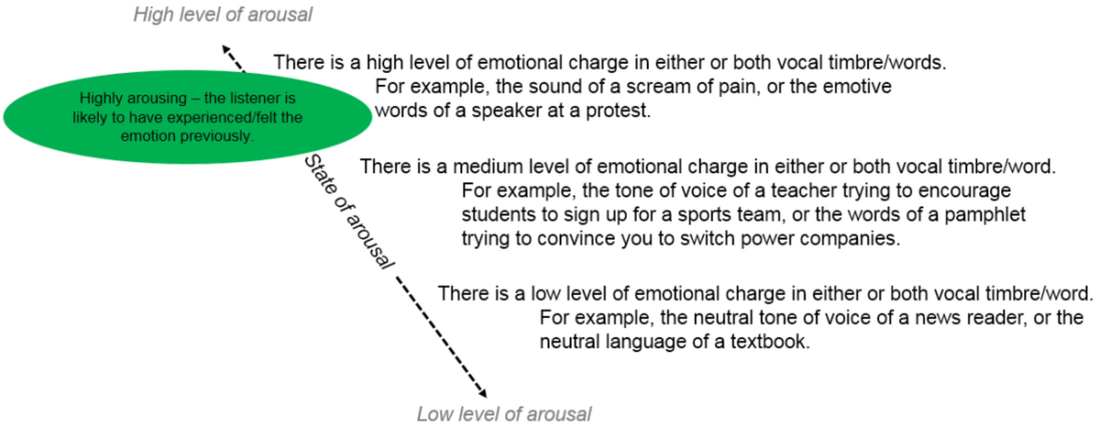

High level of arousal
Highly arousing – the listener is likely to have experienced/felt the emotion previously.
There is a high level of emotional charge in either or both vocal timbre/words.
For example, the sound of a scream of pain, or the emotive words of a speaker at a protest.
State of arousal
There is a medium level of emotional charge in either or both vocal timbre/word.
For example, the tone of voice of a teacher trying to encourage students to sign up for a sports team, or the words of a pamphlet trying to convince you to switch power companies.
There is a low level of emotional charge in either or both vocal timbre/word.
For example, the neutral tone of voice of a news reader, or the neutral language of a textbook.
Low level of arousal

Figure 8.13. Attachment in verse three of "Somebody" (De Backer, 2011, track 3). The highlighted sections show the likely level of arousal evoked by the emotive content of vocal timbre and lyrics.

To summarise the song thus far, in the local contexts (i.e., as self-contained sections), the Gotye Solo and Kimbra Solo sections produce cohesive emotional messages (i.e., the emotion conveyed in vocal timbre is mostly aligned with that conveyed in lyrics). In the global context, however, this is not necessarily the case. At the end of the first chorus (2’17”), one was left with a feeling of empathy for Gotye’s character (he was betrayed by Kimbra’s character). However, by the end of the third verse (3’02”), it has been revealed that Kimbra’s character may have had good reason to cut Gotye’s out of her life (to remove herself from an undesirable situation). This message is reinforced in the vocal timbre—Gotye’s vocal timbre evolves from sadness to anger, contrasted with Kimbra’s vocal timbre evolving from vulnerability to anger. The overarching emotional message of the song up to 3’02’’ has become unsettled, ambiguous, and convoluted—we are left unsure of whose side to take, and, therefore, of which emotional message is the “truth”.

8.2.2.3 Vocal duet: Final chorus and outro—3’03’’–4’04”.

As discussed in the preparation section above, analysing vocal timbre and lyrics for this section must be done with care. This is because the sound of the vocal line seems to take precedence over the lyrics (because the lyrics include nonwords and the vocal texture is dense, see section 8.2.1 for a more detailed discussion). It is also difficult to capture the timbres of multiple voices in the audio-visual example (AV1). This is because AV1 can only represent one set of vocal timbre features and one spectrograph at a time (the potential to develop the format of these audiovisual descriptions is discussed in section 9.2.2). The isolated vocal track used for AV1 also does not always contain all the vocal parts. Therefore, although AV1 will still be referred to in this discussion of the final chorus and outro, this section should also be read in conjunction with the full recording.

There are several interesting observations to be made in relation to the vocal line/s in this section. The first is that Kimbra is, for the most part, linguistically silent but vocally powerful in the Vocal Duet section. At the moment when her verse should lead to the chorus (3'03''), Gotye takes over and sings it instead. Kimbra's linguistic content is reduced to non-words:

> (Gotye) But you didn't have to cut me off.
> (Kimbra: [harmonising] ow)
>
> (Gotye) Make out like it never happened and that we were nothing.
> (Kimbra: [harmonising] arr)
>
> (Gotye) And I don't even need your love,
> (Kimbra: [harmonising] arr)
>
> (Gotye) But you treat me like a stranger and that feels so rough.
> (Kimbra: [harmonising] nnn)

(Gotye) No, you didn't have to stoop so low.
(Kimbra: [harmonising] arr)

(Gotye) Have your friends collect your records and then change your .number
(Kimbra: [harmonising] nnn)

(Gotye) I guess that I don't need that though.
(Kimbra: [harmonising] arr)

(Gotye) Now you're just somebody that I used to know,
(Gotye) Somebody
(Gotye) (I used to know)
(Gotye) (Somebody) Now you're just somebody that I used to know
(Gotye) Somebody
(Gotye) (I used to know)
(Gotye) (Somebody) Now you're just somebody that I used to know
(Gotye) [harmonising] I used to know, that I used to know, I used to know

(Gotye and Kimbra) Somebody.

(De Backer, 2011, track 3)

One reason for this may be that it was Kimbra's character who cut Gotye off, therefore it would be odd for her to sing this particular chorus. However, Kimbra's vocalisations on the nonwords do make it sound as though she would have said more, if it wasn't for Gotye's character's interruption. And indeed, despite being deprived of language, Kimbra's use of vocal timbre does subvert Gotye's power here. She may have lost her words, but she has not lost her sound.

Figure 8.14 shows an annotation of the backing vocals, which are led by Kimbra, from 3'03"–3'28". If one were to look only at the lyrics, then it would appear as if Gotye's character is in control at this point, delivering this chorus in much the same way as his previous one (see AV1, first chorus 1'34"–2'17", second chorus 3'03"–3'28"). However, viewing these lyrics in the context of Kimbra's vocalisations reveals a different story. Kimbra's vocalisations at 3'03", 3'10", 3'17" and 3'25" are tense, loud, and high, allowing them to penetrate through Gotye's lyrics. These vocalisations are then contrasted with those at 3'06", 3'14", and 3'21", which tend to be softer, lower, and laxer, resulting in them having the effect of *sitting under* Gotye's lyrics. The stronger vocalisations coincide with lines one (3'03"), three (3'11"), five (3'18"), and seven (3'24") in the chorus and the weaker ones with lines two (3'07"), four (3'13"), and six (3'22") such that the strong and weak vocalisations alternate:

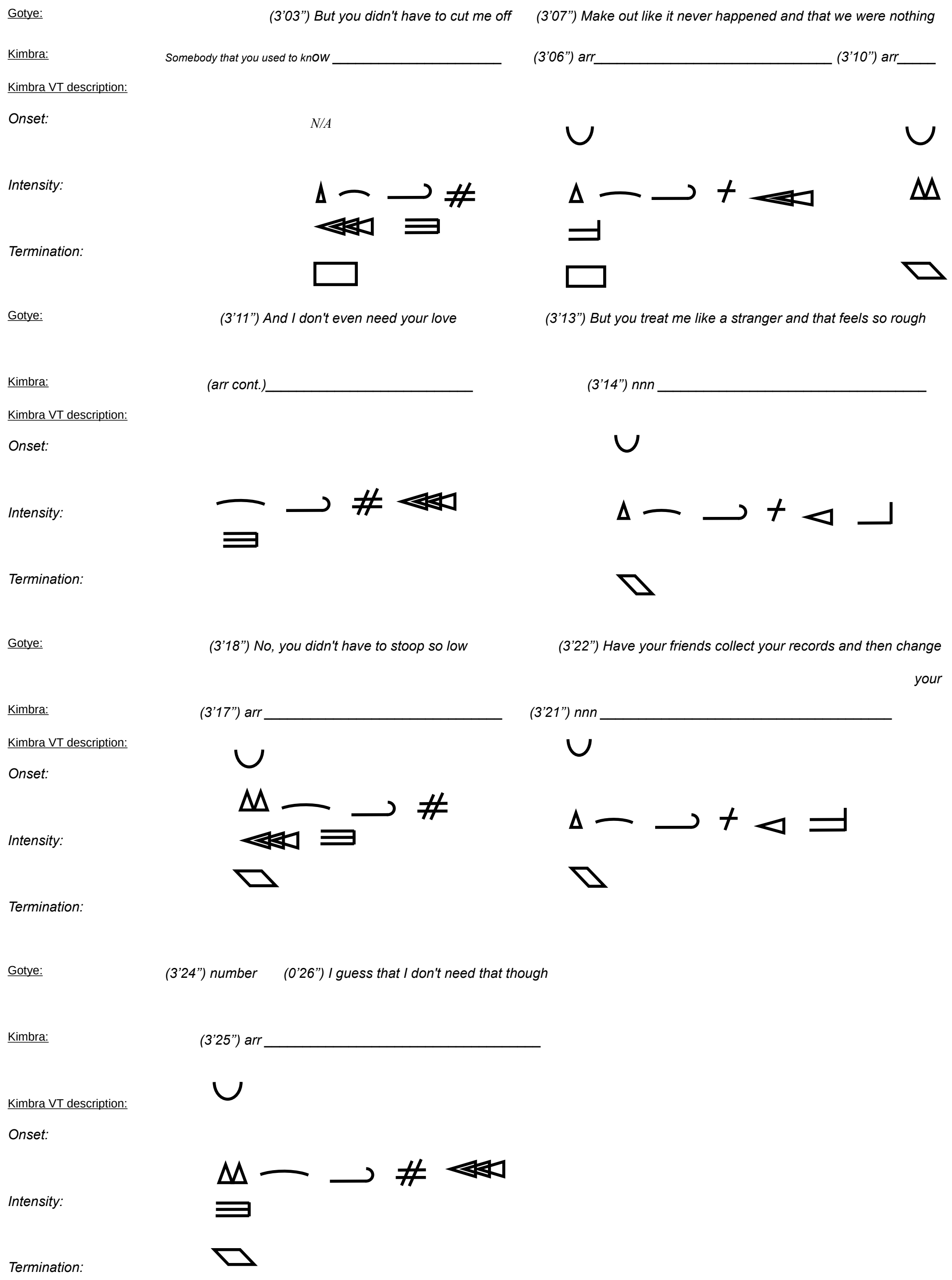
Line one = strong vocalisation, line two = weak vocalisation;

Line three = strong vocalisation, line four = weak vocalisation;

Line five = strong vocalisation, line six = weak vocalisation;

Line seven = strong vocalisation.

In this way, the seven lines of the chorus, while delivered by Gotye, are *shaped* into four phrases by Kimbra. This shaping is emphasised further by Kimbra's vocalisations. With the exception of the vocalisation at 3'14", almost always entering slightly before Gotye (see Figure 8.14)—giving the impression that Kimbra is anticipating and directing the chorus, and further emphasising her voice. Although Kimbra's vocals do not necessarily overpower the chorus, they are intrusive and certainly do compete with Gotye's lyrical message.



Gotye:
(3'03") But you didn't have to cut me off
(3'07") Make out like it never happened and that we were nothing
Kimbra:
Somebody that you used to knOW ______________________
(3'06") arr______________________________
(3'10") arr_____
Kimbra VT description:
Onset:
N/A
Intensity:
Termination:
Gotye:
(3'11") And I don't even need your love
(3'13") But you treat me like a stranger and that feels so rough
Kimbra:
(arr cont.)__________________________
(3'14") nnn __________________________________
Kimbra VT description:
Onset:
Intensity:
Termination:
Gotye:
(3'18") No, you didn't have to stoop so low
(3'22") Have your friends collect your records and then change your
Kimbra:
(3'17") arr ______________________________
(3'21") nnn _____________________________________
Kimbra VT description:
Onset:
Intensity:
Termination:
Gotye:
(3'24") number
(0'26") I guess that I don't need that though
Kimbra:
(3'25") arr ___________________________________
Kimbra VT description:
Onset:
Intensity:
Termination:

Figure 8.14. Annotated vocal features for Kimbra's vocalisations in the Duet section of "Somebody" (3'03" – 3'28"). The single quotation mark (') represents minute/s and the double quotation mark ("') represents second/s.

The second observation I wish to make in regards to this section is related to the vocal timbre's emotional content (as opposed to its general pervasiveness as discussed above). High, tense, and loud vocal timbres are a feature of Kimbra's and Gotye's vocal lines in this section. This can be seen in Figure 8.14, and also in AV1[22] which depicts Gotye's vocal timbre in this section. The tension, volume, and register here has a similar effect to that in the first two sections—it creates a feeling of almost unrelenting frustration. There are two vocal timbre features which, I believe, serve to enhance this feeling: the general lack of vibrato, and the moments of respite from the frustration.

In this section, both voices generally have tense, loud, and high vocal timbres, however they do not have a great deal of vibrato. This is despite there being plenty of opportunity for the use of such a vocal feature (the long "arr"s and "nn"s for example in 3'03" – 3'28", AV1). While this is especially true in the duet only section, it is an observation that holds true for most of "Somebody". In areas in which increasing vibrato might be used, tension and

[22] Again, it should be noted here that it is difficult to annotate multiple voices in the video descriptions. This is because, when multiple voices are present, the sound can become too dense for the spectrographs to provide accurate representation. Also, the video layout only allows for one set of Vocal Timbre Features to be displayed at a time. Annotating two voices on the one set of features, or adding another set of features, could result in the annotation becoming cluttered and difficult to follow. Refining the format of the video description to allow for the annotation of multiple voices at once should be explored in further research.

roughness are instead employed. In this way, it is as though this tension and roughness are being used as a substitute for vibrato.

One example of this can be found at 3'18"–3'21" with the lyrics "You didn't have to stoop so low" (see AV1). The spectrographs for these lyrics can be seen in Figure 8.15. Here, darker harmonics can be seen from halfway through the word "stoop" to the end of the phrase. These darker harmonics are a visual representation of the noise that can be heard within the vocal timbre at this point. This noise is created by the increased tension, volume, and roughness in the voice. Of course, one would expect to see denser harmonics in this spectrograph as it depicts both Gotye's voice as well as additional vocalisations. However, these harmonics remain dense, and even increase in density, even after the additional vocalisations have dropped out. This supports what is heard aurally: Gotye's vocal timbre becomes significantly noisier (rougher, tenser, louder, and higher) at the end of this line. Substituting vibrato for increased roughness, tension, register, and volume in this way creates a feeling of rawness—there is no relief for the listener from the tension in the vocal timbre, just as there is no relief for the singers from the emotions of frustration and anger.

Darker partials can be seen here due to Gotye's loud and tense vocal timbre occurring at the same times as the loud, tense vocal timbre in the vocalization. However, these partials become noticeably darker halfway through the word "stoop".

The dark, dense partials indicate a louder volume and increasing levels of noise (created by increased tension and roughness). This persists even after the vocalizations have dropped out, suggesting that this noise exists in both voices.

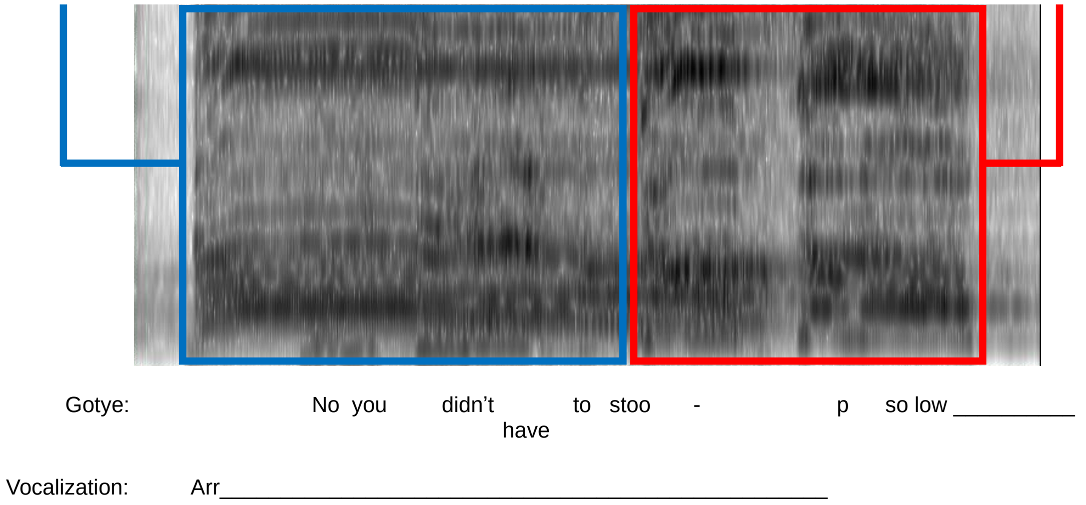


Figure 8.15. Spectrograph of the line "No, you didn't have to stoop so low" (De Backer, 2011, track 3, 3'18"–3'21"). Increasing levels of noise, created by increasing roughness, tension, and volume, can be seen towards the end of this line.

While there may not be any extended periods of relief from this tension, there are moments of respite. The first comes in Gotye's aspirate onsets at 3'23" with the word "records" (see AV1). As in the first chorus (see discussion in 8.2.2.1), there are other aspirate onsets present here, however the one occurring with "records" is particularly significant as it is not linguistically necessary (i.e., the word *records* can be pronounced without an aspirate onset; e.g., see glossary in Collins & Mees, 2013, pp. 295–308). When this onset was heard in the first chorus (1'53"), it evoked a sense of sadness and volatility. It is likely that the same sense of volatility is evoked here as this onset again occurs quickly and without warning, once more indicating that Gotye's character is only just in control of his emotions.

However, it may not necessarily be that the sense of sadness evoked in the first chorus holds true in the present chorus. Here, it may be that this aspirate onset is heard through the lens of Kimbra's breathy onsets in verse three. At that time, aspiration was indicative of vulnerability (see discussion of Kimbra's vocal timbre section 8.2.2.2). Given that aspiration was such a prevalent feature in verse three, and given that Gotye's aspirate onset at 3'23" occurs only a short time after verse three, it is possible that Gotye's onset also evokes a sense of vulnerability.

Moments of respite also occur at 3'34", 3'41", 3'44", 3'48", 3'52", and 3'56" (see AV1). At these points, the words "I used to know" are delivered with a softer, breathier, vocal timbre reminiscent of that heard in verses one (0'19") and three (2'33") (see AV1). The sudden reversion back to this timbre brings with it the memory of emptiness, pain, and vulnerability (as they are the emotions initially conveyed by this kind of timbre). These softer phrases occur only intermittently between Gotye's otherwise still-tense lyrics. However, the contrast between these softer interjections and the still-tense lyrics creates the feeling that the anger is beginning to burn itself out, and the sadness is beginning to set in in earnest.

The emotional message of this vocal timbre is aligned with that conveyed by the lyrics. The words "I used to know" express a sense of loss. This loss is all the more acute as the lyrics do not speak of something that he used to have, but something that he used to know. It is not an object that is gone, but a part of the singer's life and, possibly, a part of their identity. On the final word, "somebody" (3'58''), the anger finally gives way to sadness and emptiness. The timbre is breathy and soft, and the lyric is detached (speaking of a disembodied "somebody") (see AV1).

8.2.2.4 Concluding thoughts.

Emotional themes

Throughout the analysis, primary emotional themes and their nuances have been identified within "Somebody". These have been gathered in the Emotional Map in Figure 8.16. The primary emotional themes of anger and sadness have been identified based on both lyrics and vocal timbre. Betrayal is a nuance common to both anger and sadness, and this is mapped in Figure 8.16. The Emotional Map provides a form of diagrammatically documenting the emotions of "Somebody" in a concise way.

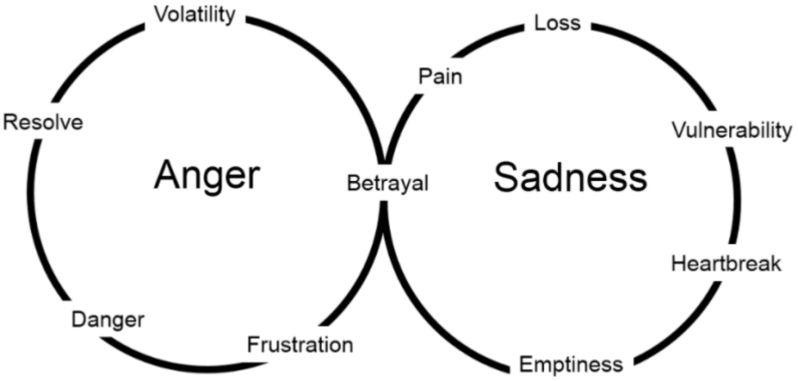


Figure 8.16. Emotional Map of "Somebody" (De Backer, 2011, track 3). This vocabulary set lays out the primary emotional themes and their nuances and variations.

Final thoughts

Up until the end of verse three (3'02"), the emotional message in the vocal timbre and lyrics evoke an overarching emotion of anger, and frustration. While each character may not necessarily begin here (Gotye's begins with sadness, Kimbra's with vulnerability), it is to these emotions that each eventually arrive. The lyrics reveal that these emotions are elicited in each singer for different reasons. Kimbra's character is angered/frustrated by the way she was treated, which is evidenced by her singing primarily about the relationship and Gotye's character's actions while they were together (e.g., "Now and then I think of all the times you screwed me over" (2'33"), and "But I don't want to live that way, giving into every word you say" (2'48") (see AV1)). Gotye's character is angered/frustrated by the way he is now being treated, which is evidenced by him singing primarily about how the relationship ended and Kimbra's character's actions since the breakup (e.g., "You didn't have to cut me off" (3'03"), and "I don't even need your love, but you treat me like a stranger and it feels so rough" (3'11") (see AV1)).

The lyrics "somebody that I used to know" further emphasise how both characters arrive at anger from different starting points. At the end of the third verse, both singers sing "somebody that I used to know" with anger and frustration (Kimbra at 3'00", Gotye at 1'59", see AV1). However, Kimbra's character is angry that, throughout their relationship, Gotye's character was hung-up on somebody that he used to know. This is illustrated by the lyrics "You said that you could let it go, and I wouldn't catch you hung up on somebody that you used to know", which suggest that Gotye's character was hung up on an old lover, or preoccupied with an idea of who Kimbra was (perhaps she is a different person now than when she entered the relationship, and it is this previous version of herself that is the "somebody" that Gotye's character used to know). Gotye's character, on the other hand, is frustrated and angered by the way Kimbra's has cut him out of her life, effectively making her just somebody that he used to know.

At the beginning of the vocal duet (3'03"), another element is added to this anger and frustration—power. By cutting Kimbra off at the end of the third verse, it is as though Gotye's character is asserting his power in this situation. However, Kimbra does nonetheless maintain her own voice, using nonword vocalisations to direct the phrase structure of the lyrics (3'03" – 3'28"). It may be that Kimbra's character has also asserted her power in other ways. By making "out like it never happened" (1'37") and acting as though they "were nothing" (1'39"), Kimbra's character is refusing to recognise that the relationship existed, perhaps in the same way that Gotye's refused to recognise Kimbra's character's autonomy while she was in the relationship (as she expresses through "But I don't want to live that way, giving into every word you say" (2'48")) (see AV1).

Through the process of considering how emotion conveyed in vocal timbre impacts emotional perception of lyrics, this study has shown that the interplay between vocal timbre and words form complex relationships in "Somebody". This is a multilayered song in which the singers express a range of emotions through the manipulation of the relationship between vocal timbre and lyrics. Even if, at the song's conclusion, we are left still wondering at what the true emotional message may be, we are certainly in no doubt that there is no "winner" in this situation—both characters have hurt and been hurt.

8.3 Adele's "Someone Like You"

"Someone Like You" is a popular vocal song performed by Adele Adkins, who performs as "Adele". Released on January 24, 2011 "Someone Like You" forms part of Adele's second album titled *21* (Adkins & Wilson, 2011, track 11). In 2012, the song received the Grammy award for Best Pop Solo Performance, and has been voted, in conjunction with the Official Charts Company's 60th anniversary, the third most favourite single of the past 60 years. “Someone Like You”—hereafter “Someone” (Adkins & Wilson, 2011)—has sold over 1,800,000 copies in the United Kingdom, 490,000 in Australia, and 6,000,000 in America. These strong sales figures demonstrate that the song appeals to a variety of listeners across the world. One reason for this is that, like “Somebody” discussed in section 8.2, “Someone” explores themes that are common to many people’s lives—rejection, love, loss, and pain.

8.3.1 Preparation.

The recording

The recording to be used as the basis for this study is:

Adkins, A. & Wilson, D. (2011). Someone Like You [Recorded by Adele]. On *21* [CD, digital download]. London, UK: XL.

Descriptive tools

Audiovisual attachment 2 (AV2 for short) shows the audiovisual description of vocal timbre in “Someone”. As with the above analysis of “Somebody”, this video description synthesises several different descriptive tools: syncing lyrics and recording with spectrographs, annotated Vocal Timbre Features, and identification of points of audible breath. The result of combining these tools is an annotated, audiovisual, video description of Adele's vocal timbre.

Analytical structure

I have chosen to structure the present analysis according to key sections within the song. Based on the role of the voice in "Someone", two main sections have been identified*: Instrumental Only*, and *Adele Solo* (a subsection of which is the *Bridge*, which includes vocal layering) (see Figure 8.17). The Instrumental Only sections are characterised by the absence of the vocal line, and the presence of the solo piano. Vocal timbre and lyrics cannot, of course, be analysed in the Instrumental Only sections as these sections do not include the vocal line.

The Adele Solo sections are characterised by the presence of the vocal line, with the piano playing an accompanying role. Adele's voice is placed in front of the piano, something which is particularly noticeable at entry points (e.g. 0'13", 1'50", 3'23")[23] where the piano drops back in dynamic (i.e., volume) to give the vocals *space*. This serves to emphasise the vocal line in these sections.

[23] In the times presented in this analysis, the single quotation mark (') represents minute/s and the double quotation mark (") represents second/s.

Other qualities also serve to emphasize the vocal line. For example, in the Adele Only sections, the vocal line is noticeably louder than the piano accompaniment. This dynamic contrast is especially noticeable in the chorus, where Adele's increase in volume is emphasised by her change in register. Further, Adele's vocal line is rather varied. This variation is emphasised by the repetitive nature of the piano accompaniment (for example, the opening piano solo (0'00"–0'12") is repeated verbatim when Adele enters (at 0'13")) and, therefore, serves to draw attention to the vocal timbre and lyrics in these sections. Adele, in effect, does not stop singing for a significant amount of time from 0'13"–4'29". While there are short instrumental interludes (e.g., 1'12"–1'13", 1'48"–1'51", 2'34"–2'37", 3'22"–3'24"), these are brief and repetitive (usually continuing a musical idea from a previous phrase, or beginning a new idea for the next phrase, but never constituting a phrase within themselves). This also helps to maintain the listener's attention as it provides the listener with a continual stream of text for one to listen to—in other words, the story does not stop.

I have identified the Bridge as a subsection of the Adele Solo section as the Bridge is characterised by the presence of *vocal layering,* with Adele's solo line supported by piano accompaniment and voices singing in harmony. The additional layers of vocals present at this point have the effect of creating a denser sound, thus foregrounding the *texture* and the sound of the voices above all else. This being said, despite this thicker vocal texture, the backing vocals do not actually compete with Adele's solo line. Rather, these vocals sit behind the solo line—allowing them to support, and to build a sense of drama, without overpowering. As the solo line is not eclipsed by these backing vocals, and as the lyrics continue to present variation and new lyrical material, analysing the synergies between vocal timbre and lyrics is possible—for this reason the Bridge will be analysed along with the Adele Solo section more generally. However, care must be taken in this section due to the emphasis on vocal texture created through the layering of voices (for which reason I have felt it necessary to make this distinction here).

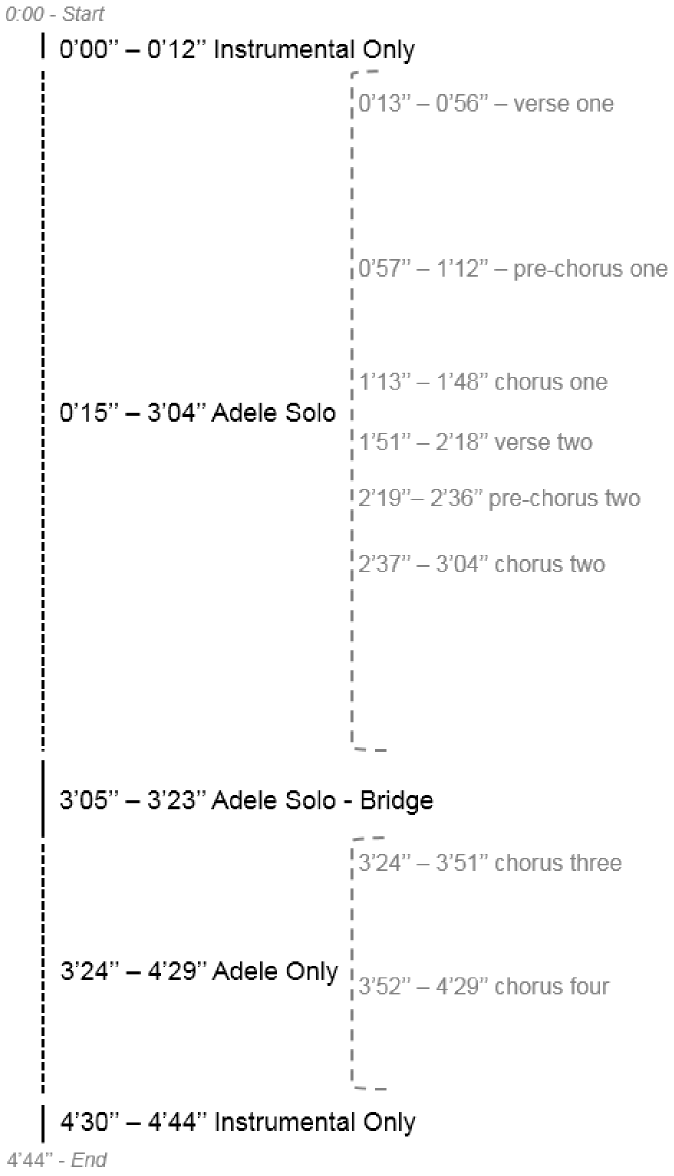

0:00 - Start
0'00" – 0'12" Instrumental Only
0'13" – 0'56" – verse one
0'57" – 1'12" – pre-chorus one
1'13" – 1'48" chorus one
0'15" – 3'04" Adele Solo
1'51" – 2'18" verse two
2'19"– 2'36" pre-chorus two
2'37" – 3'04" chorus two
3'05" – 3'23" Adele Solo - Bridge
3'24" – 3'51" chorus three
3'24" – 4'29" Adele Only
3'52" – 4'29" chorus four
4'30" – 4'44" Instrumental Only
4'44" - End

Figure 8.17. Timeline of "Someone" (Adkins & Wilson, 2011, track 11) showing the different sections of the song (Instrumental Only, Adele Solo, Adele Solo—Bridge) and when they occur. The single quotation mark (') represents minute/s and the double quotation mark (") represents second/s.

8.3.2 Analysis.

Unlike the analysis of “Somebody” in 8.2.2, which was chronological in structure, here, I will analyse all the verses of “Someone”, then all the prechoruses and Bridge, and finally all of the choruses (see Figure 8.17 for a timeline of these events). This structure has been chosen because, as will be discussed in the following sections of this analysis, each of these three sections (verse, prechorus/bridge, and chorus) expresses distinct, but related, emotional messages. The times given for specific events are in reference to the video description AV2. One should read this analysis in conjunction with AV2.

As with the analysis of “Somebody” in section 8.2, as emotional themes and nuances are identified throughout this analysis, they will be included in the Emotional Map (Figure 8.29). This mode of presentation provides a succinct, visual way to present the emotional themes and nuances identified within the song.

8.3.2.1 Verses one and two, 0’13’’–0’56’’ and 1’51’’–2’18’’.

Overall, Adele’s moderate use of roughness, breath, and vibrato in verses one and two gives the impression that her character is in control of her emotions. For example, after the aspirate onset at the opening of verse one (see AV2, 0’13’’, this onset will be discussed in more detail in the following paragraphs), Adele's vocal timbre is smooth and clear (see AV2, 0'13"–0'56"). A "comfortable" amount of vibrato is used. That is, Adele's vocal timbre neither employs a lot of vibrato (which may be more indicative of a louder, shouted, sound) nor is vibrato sparse (which may be indicative of disparity and despair). Rather, Adele uses a moderate amount of vibrato, making her vocal timbre seem ordinary, and not overly emotive. Similarly, the ends of Adele's words (terminations) are neither overly articulated (which might suggest that Adele is shouting, or forcing the words out), nor too tapered (which might be a symbol of disheartenment—she does not have enough energy to finish her words). Instead, these terminations are controlled, giving the impression of a neutral, stable emotional state. The same kind of reserved, moderate, ordinary treatment of these vocal features can be seen in verse two (see AV2, 1'51"–2'18"), which reinforces this impression that, across, the two verses, Adele's character is in command of her emotions.

As verse one develops, however, the listener begins to hear cues to suggest that this impression of emotional control is an illusion. Nuances in her vocal timbre give the listener clues that the character's emotional state is not as stable as might be supposed (based on the message projected through the lyrics). Rather, there exists an underlying emotional turmoil which is projected in Adele's vocal timbre in several ways.

The first is the contrast between the tense/lax, high/low, and strong/weak features of Adele's vocal timbre in the first verse. Here, a strong–weak pattern is created through Adele's vocal timbre (see Figure 8.18). The strong sections are measured and restrained, and are therefore in alignment with the neutral emotional message evoked by the use of vibrato, breath, and roughness. They suggest that Adele's character is in control. The weaker sections, however, create a sense of sadness (which is a primary emotional theme and is mapped in Figure 8.29). They are less resonant, and therefore, sound less energetic. It is as though the character's energy is being spent *processing* the break up, there is little left for her to project her sound.

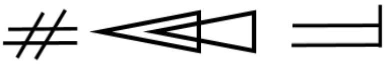
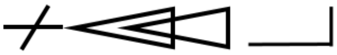
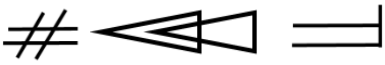
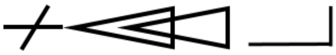
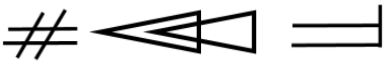
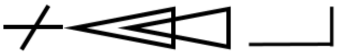
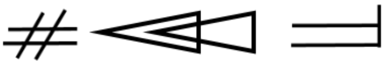
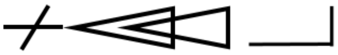
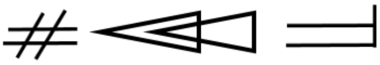
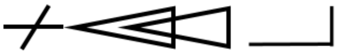
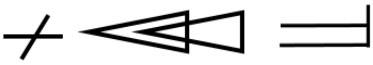
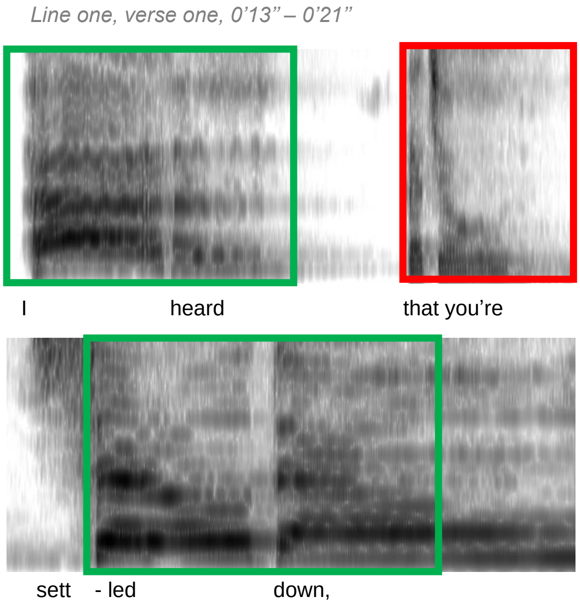
Vocal timbre features		
Lyric	*I heard*	*that you're*
Vocal timbre features		
Lyric	*settled down*	*That you*
Vocal timbre features		
Lyric	*found a girl*	*and you're*
Vocal timbre features		
Lyric	*married now. I heard that your dreams came true.*	*Guess she gave you things I didn't give to you.*
Vocal timbre features		
Lyric	*Old friend, why are you so shy?*	*Ain't like you to hold back*
Vocal timbre features		
Lyric	*or hide from the light.*	

Figure 8.18. Vocal contrasts highlighted through the Vocal Timbre Features. The tense/lax, high/low, and strong/weak features of Adele's vocal timbre in the first verse (see AV1; Adkins & Wilson, 2011, track 11).

Figure 8.19 shows the spectrographs of the opening line of “Someone”: “I heard that you’re settled down”. These spectrographs show a small section of the strong–weak pattern outlined in Figure 8.18. The denser harmonics on the words “I heard” and “settled down” visually confirm what is heard aurally—that these words are sung with a tenser, louder, and higher vocal timbre. By contrast, a weak vocal timbre is demonstrated on the words “that you’re”. Here one can see more space in the spectrograph (represented by the white patches) which indicates that the harmonics at this point are less dense—something which can be indicative of a softer, lower, more lax sound.



"I heard" and "settled down" are both sung in a tenser, higher, and louder vocal timbre than that used for the words "that you're".
This is shown by the darker, denser partials shown throughout the spectrum (shown in green boxed).
The weaker vocal timbre on "that you're" is shown through the lighter, less dense partials on these words (shown in the red box).

Figure 8.19. Spectrograph of line one, verse one, showing the strong/weak contrasts present within the vocal timbre. The single quotation mark (') represents minute/s and the double quotation mark (") represents second/s.

Neither the emotional state evoked by the strong nor the weak vocal timbre is retained for long, however. Rather, Adele moves swiftly between the two timbres/emotional states (see Figure 8.18 for more detail on when Adele moves from strong to weak). That Adele does not linger long in either emotive state suggests that the neutrality established through the vibrato, breath, and roughness is not reliable. Instead, Adele's character's underlying emotional state is much more unpredictable, moving without warning from a sense of resolve, to a sense of sadness.

In verse one, onsets provide another clue to suggest that Adele's character is not emotionally stable. The very first sound one hears when Adele enters is an aspirated onset on the word "I" (see AV2, 0'13"). This aspirated onset is noisy, which (as discussed in Chapter 7, and in Part II of this thesis) may suggest negative emotions. The noise produced from this onset is also soft. This creates a sense of closeness (one needs to be close to hear such soft sounds), and may therefore suggest that the emotion expressed by Adele's character here is not only negative, but is very personal. That is, it is an emotion that one does not project out into the world, but rather expresses quietly to oneself. In this way, this aspiration may evoke feelings of sadness. Later, this feeling is validated and intensified when heard in the context of the tense/lax, loud/soft, high/low contrasts in verse one.

Verse one		Prechorus one		Verse two		Prechorus two	
Time	Onset	Time	Onset	Time	Onset	Time	Onset
0’13”	Aspirate Onset	0’56”	Aspirate Onset	No instances of “I”.		2’18”	Creak Onset
0’29”	Creak Onset	1’00”	Simultaneous Onset			2’22”	Simultaneous Onset
0’39”	Creak Onset	1’01” (I’d)	Simultaneous Onset			2’24”	Simultaneous Onset
		1’03”	Simultaneous Onset			2’25” (I’d)	Simultaneous Onset

Figure 8.20. The treatment of the word “I” throughout verse one and two, and prechorus one and two. The single quotation mark (‘) represents minute/s and the double quotation mark (“) represents second/s.

The next two presentations of “I”, at 0’29” and 0’39”, are sung with a creak onset (see AV2, and see Figure 8.20). Figure 8.21 shows the contrast between these creak onsets and the aspirated onset at 0’13”. The onset at 0’13” begins much weaker. Air in the voice can be seen in the presence of less dense harmonics in the initial part of the note. These harmonics gradually become denser and a strong sense of pitch also gradually begins to take shape. By contrast, the onsets at 0’29” and 0’39” present a very different spectrograph. These creak onsets feature louder noise, which is depicted by the presence of distinguishable harmonics in the upper spectrum. These harmonics, while not as dense as those in the lower spectrum where the pitch sits and therefore where a higher sustain of harmonics should be seen, are noticeably more distinct and regular than those seen in the aspirate onset. This increased volume in the glottal onsets may also symbolise more overt emotional states. The sense of sadness evoked 12 seconds earlier by the softer aspirated onset is stripped away, and instead a sense of harsh and more jarring negative emotion may be evoked. One such emotion which fits with this context of sadness is pain (this relationship is mapped in Figure 8.29). On these onsets of "I", Adele's character sounds pained—perhaps by the knowledge that her ex-boyfriend has moved on ("found a girl"), is content ("dreams came true"), and is fulfilled by not being with her ("she gave you things, I didn't give to you") (see AV2).

The “I” sung at 0’13’’ has an aspirate onset. This can be seen by the partials slowly spreading through the spectrum, beginning weaker, and becoming more dense (shown in the green box). A strong sense of pitch does not occur until later in the note (indicated in the blue box).

The “I”s sung at 0’29’’ and 0’39’’ have a creak onset. This can be seen by the presence of noise (indicated by lighter, but distinct partials) above the note (shown in the red box).

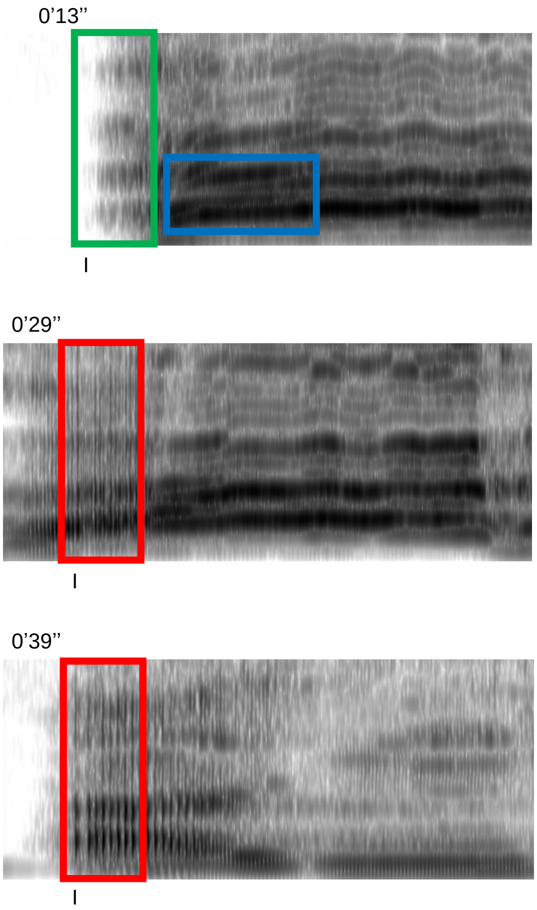


Figure 8.21. Spectrographs of the three “I”s present in verse one. The first “I” features an aspirate onset, the second two both feature creak onsets. The single quotation mark (‘) represents minute/s and the double quotation mark (“) represents second/s.

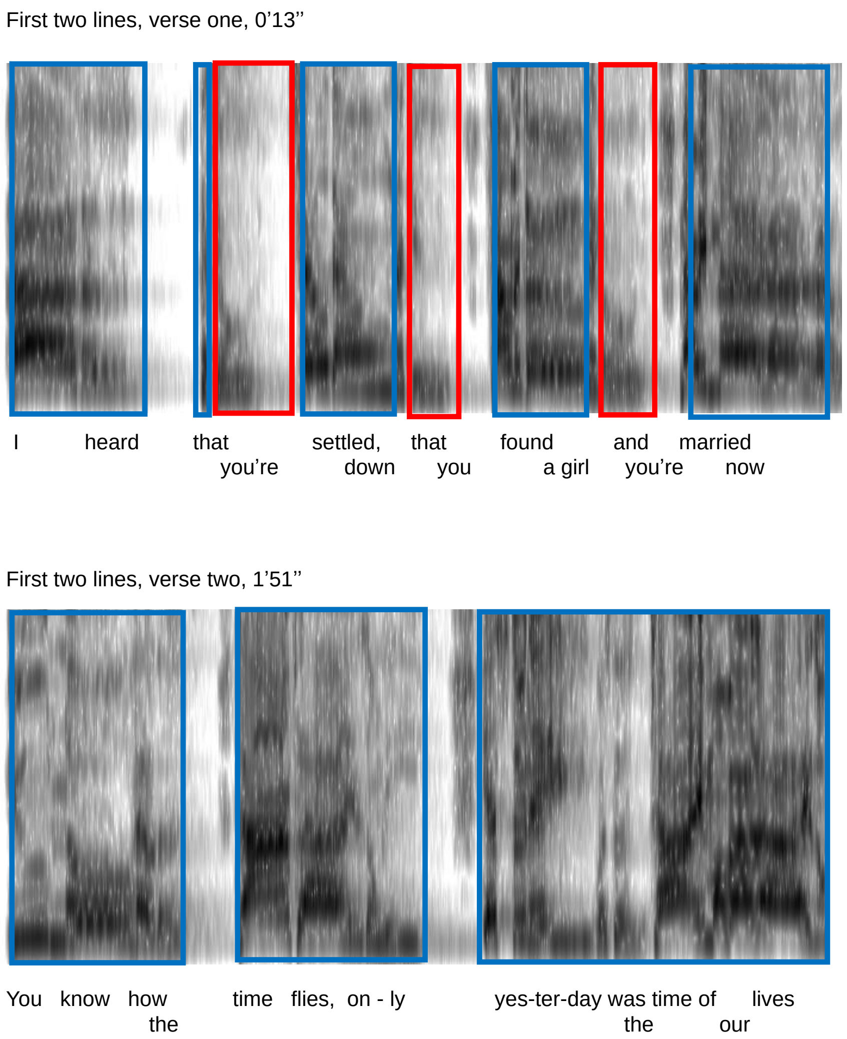
It is perhaps significant here that, in the first verse, the word "I" is always sung with either an aspirate or creak onset. This treatment of "I" means that it is always imbued with a negative emotion (whether sadness, or pain). Given that "I" is a word used to refer to oneself (a concept that will be revisited in the analysis of the choruses in 8.3.2.3), singing this word with such negative onsets may symbolise Adele's character's emotional state—she is saddened and pained, so her *personal pronoun* is saddened and pained too. This sense of pain is prolonged throughout the first verse by the recurrent use of creak and aspirate onsets (see Figure 8.22) on the words "old", "why", and "ain't", as in the lyrics:

> Old friend, why are you so shy?
> Ain't like you to hold back or hide from the light. (Adkins & Wilson, 2011, track 11).

Verse one		Prechorus one		Verse two		Prechorus two	
Time	Onset	Time	Onset	Time	Onset	Time	Onset
0’13’’ (I)	Aspirate Onset	0’57’’ (I)	Aspirate Onset	2’09’’ (summer)	Aspirate Onset	2’19’’ (I)	Creak Onset
0’19’’ (settled)	Aspirate Onset			2’13’’ (surprise)	Aspirate Onset	2’23’’ (stay)	Aspirate Onset
0’23’’ (found)	Aspirate Onset			2’14’’ (of)	Creak Onset		
0’29’’ (I)	Creak Onset						
0’36’’ (she)	Aspirate Onset						
0’39’’ (I)	Creak Onset						
0’43’’ (old)	Creak Onset						
0’44’’ (friend)	Aspirate Onset						
0’45’’ (why)	Creak Onset						
0’46’’ – 0’47’’ (so shy)	Aspirate Onset						
0’49’’ (ain’t)	Creak Onset						

Figure 8.22. Creak and aspirate onsets in verses one and two, and prechoruses one and two. The single quotation mark (‘) represents minute/s and the double quotation mark (“) represents second/s.

Verse two presents a very different emotional message. It features neither variations in sustain (in terms of tenseness, volume, and register) nor significant differences in onset. Rather, verse two presents a more resolved emotional message (resolve that has, perhaps, come from her grief, this relationship is mapped in Figure 8.29). They lyrics are presented in a much steadier, and stronger vocal timbre. Figure 8.23 contrasts spectrographs of the first two lines of verse one with the first two lines of verse two. Verse two shows less variation in sustain. All words are sung with roughly the same tension, volume, and register. Because these words are stronger and more resonant than words shown in red boxes in Figure 8.23, their termination is more distinguishable. This can be seen by words in verse two having more constant, darker harmonics, and clearer terminations. The resolve, which was wavering in verse one as shown by the contrasted weak—strong intensities seems to have returned in earnest in verse two—speaking in terms of sustain, we can believe that Adele's character is in control of her emotions.



First two lines, verse one, 0’13’’

First two lines, verse two, 1’51’’

Words sung with stronger intensities are shown in blue boxes, words sung with weaker intensities are shown in red boxes.

The first line shows much more contrast between weak-strong timbres than the second.

Figure 8.23. Spectrographs of the first two lines of verse one compared with the first two lines of verse two. The first verse demonstrates a greater variety of strong–weak intensities than the second. The single quotation mark (‘) represents minute/s and the double quotation mark (“) represents second/s.

Similarly, variation in onset is not seen in verse two (see Figure 8.22). Verse two features only two aspirate onsets, and one creak onset, compared to verse one which features six aspirate onsets, and five creak onsets. This again reinforces the sense of resolve present in the second verse. From 1’51”–2’18”, Adele’s character achieves a sense of emotional clarity and stability which was not present in verse one (or, as will be discussed in 8.3.2.3, in the first chorus).

Explanations for these contrasting emotional states between verses one and two may be found in the lyrics. The lyrics of the first verse are:

> I heard that you're settled down
> That you found a girl and you're married now.
> I heard that your dreams came true.
> Guess she gave you things I didn't give to you.
>
> Old friend, why are you so shy?
> Ain't like you to hold back or hide from the light. (Adkins & Wilson, 2011, track 11)

These lyrics refer to the character’s present relationship with her ex-boyfriend. In this relationship, everything is going right for the ex-boyfriend. This is indicated in the first three lines which list the positive things that have happened to the ex-boyfriend since breaking up with Adele’s character. By contrast, not much is going right for the character Adele is portraying. This is illustrated in the last three lines, which feature Adele’s character reflecting on (and potentially blaming herself for) her shortcomings ("Guess she gave you things I didn't give to you"), how she doesn't understand the person her "old friend" has become ("Ain't like you to hold back"), and how she doesn't understand how and why their relationship has changed from what it was in the past ("Old friend, why are you so shy?").

The sense of disconnect and lack of understanding present in the lyrics is reinforced by the vocal timbre, which is generally misaligned. Chronologically speaking, the first time this misalignment occurs is on instances of the word "I" (0'13'', 0'29'', 0'39''). In the first verse, "I" is sung either with an aspirate or creak onset. As discussed above, pain and sadness are in some way associated with both onsets. As every presentation of "I" is treated in this way, the result is that this word always appears with some kind of associated negative connotation. However, *I* is not necessarily a positive or negative word. Rather, it is a word which acts as the linguistic representation of the character's self and which therefore may also be considered very personal. By always steeping this word in negativity, the vocal timbre is slightly misaligned with the lyrics, which creates an unsettling effect—Why is Adele's character being so hard on herself? Why did this relationship mean so much to her? Overall, then, the emotional message is ambiguous (see Figure 8.24). Certainly, it is normal to be sad and pained after a break up, but why is she placing all the emotional burden on herself and none on the ex-boyfriend? It would be possible for the word "friend" in the first verse to be delivered with more hate and anger, however these are not emotions that Adele's character directs towards the ex-boyfriend.

Despite (or perhaps because of) this misalignment, this emotional message is likely highly arousing for the listener. Not only are relationship breakdowns a common experience (whether romantic or otherwise), but the experience of blaming oneself, and deconstructing one's own actions, would also be familiar to many listeners. In this way then, in the presentation of the word "I" in the first verse, Adele's character has delivered an emotional message that is rooted in the human experience (see Figure 8.25).

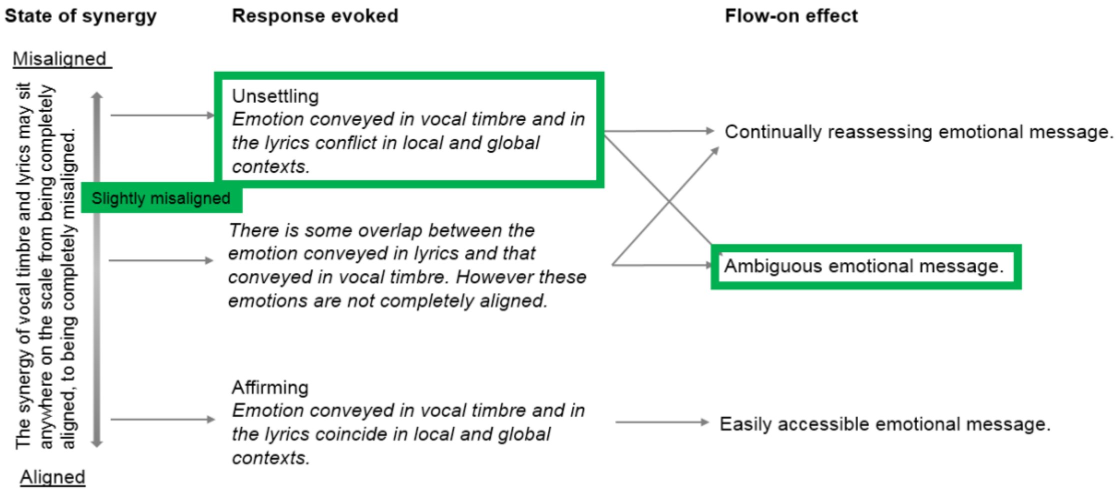

State of synergy
Response evoked
Flow-on effect
Misaligned
The synergy of vocal timbre and lyrics may sit anywhere on the scale from being completely aligned, to being completely misaligned.
Aligned
Unsettling
Emotion conveyed in vocal timbre and in the lyrics conflict in local and global contexts.
Slightly misaligned
There is some overlap between the emotion conveyed in lyrics and that conveyed in vocal timbre. However these emotions are not completely aligned.
Affirming
Emotion conveyed in vocal timbre and in the lyrics coincide in local and global contexts.
Continually reassessing emotional message.
Ambiguous emotional message.
Easily accessible emotional message.

Figure 8.24. Cohesiveness of “Someone” in the “I”s in particular, the first three lines of verse one in general, and in the Bridge and prechoruses. Highlighted sections show the synergies between vocal timbre and lyrics.

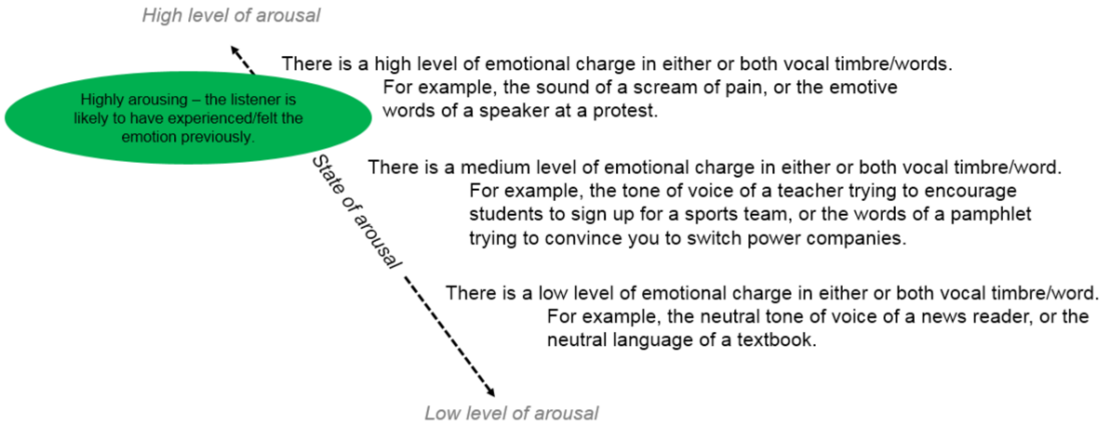

High level of arousal
Highly arousing – the listener is likely to have experienced/felt the emotion previously.
There is a high level of emotional charge in either or both vocal timbre/words. For example, the sound of a scream of pain, or the emotive words of a speaker at a protest.
State of arousal
There is a medium level of emotional charge in either or both vocal timbre/word. For example, the tone of voice of a teacher trying to encourage students to sign up for a sports team, or the words of a pamphlet trying to convince you to switch power companies.
There is a low level of emotional charge in either or both vocal timbre/word. For example, the neutral tone of voice of a news reader, or the neutral language of a textbook.
Low level of arousal

Figure 8.25. Attachment of “Someone” in the “I”s in verse one, in all of verse one more generally, and in all four choruses. The highlighted sections show the likely level of arousal evoked by the emotive content of vocal timbre and lyrics.

Indeed, the emotional message of the first three lines is misaligned with the emotion conveyed in the vocal timbre. The lyrical message here is a positive one. The person has settled down, found love, and is happy. However, the emotional cues within the vocal timbre suggest sadness and confusion (created by the contrast between strong and weak intensities). This reinforces the unsettling feeling (see Figure 8.24) established in the initial "I", and maintained in subsequent "I"s, as the listener continually reassesses the lyrics against the vocal timbre to determine why the singer is expressing these sad emotions. Is it regret over the falling out with a family member, remorse at having missed years with an estranged friend, or heartbreak over a relationship break down? (All possibilities are derivatives of the primary emotion of sadness, as depicted in Figure 8.29). At this point, who is being sung about is not certain. It is not until the fourth line "guess she gave you things I didn't give to you" that the listener can be quite sure that this is a song about an ex-boyfriend. Regardless of who is the subject of the song, the message remains highly arousing (see Figure 8.25) since, as discussed above, the experience of blaming oneself, and deconstructing one's own actions, would also be familiar to many listeners.

The last three lines of verse one offer a lyrical message which is aligned with that conveyed in the vocal timbre. The lyrics suggest confusion which gives rise to the sadness and pain expressed by Adele's character through the sound of her voice. This creates an affirming effect, we now understand that Adele's character is singing about being jilted by a lover and this is why she is sad and pained. The emotional message here is easily accessible (see Figure 8.26). Nonetheless, this emotional message remains arousing as it still speaks of a common human experience to which many listeners can relate (see Figure 8.25).

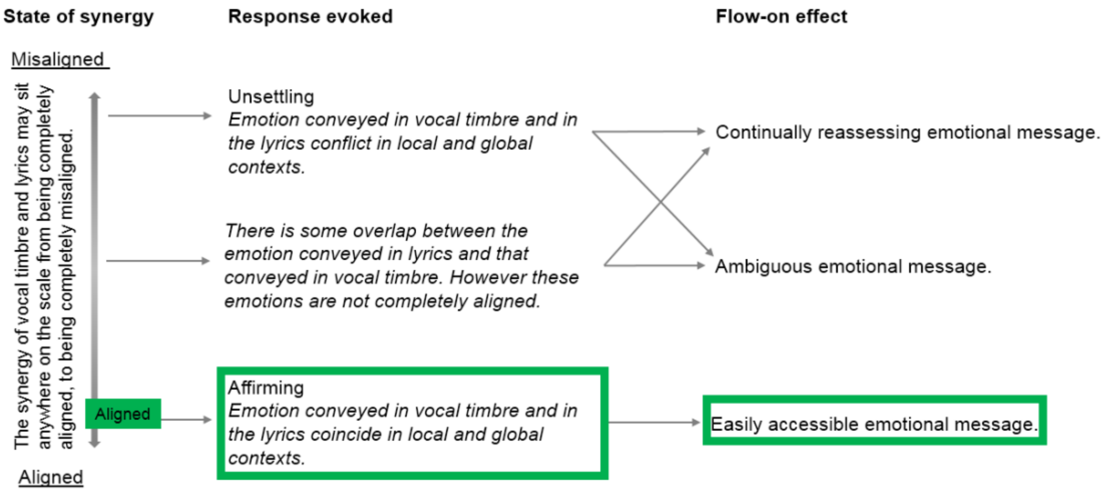

State of synergy
Response evoked
Flow-on effect
Misaligned
The synergy of vocal timbre and lyrics may sit anywhere on the scale from being completely aligned, to being completely misaligned.
Aligned
Aligned
Unsettling
Emotion conveyed in vocal timbre and in the lyrics conflict in local and global contexts.
There is some overlap between the emotion conveyed in lyrics and that conveyed in vocal timbre. However these emotions are not completely aligned.
Affirming
Emotion conveyed in vocal timbre and in the lyrics coincide in local and global contexts.
Continually reassessing emotional message.
Ambiguous emotional message.
Easily accessible emotional message.

Figure 8.26. Cohesiveness of “Someone” in the last three lines of verse one in general, in verse two, and in the choruses. Highlighted sections show the synergies between vocal timbre and lyrics.

The lyrics of the second verse convey a different emotional message. They refer to the character’s past relationship with her ex-boyfriend:

You know how the time flies
Only yesterday was the time of our lives
We were born and raised
In a summer haze

> Bound by the surprise of our glory days. (Adkins & Wilson, 2011, track 11)

Adele’s character is much happier “living in the past” as this is the kind of relationship she understands. Reflecting on their “glory days” makes her feel comfortable and secure, and so she sounds it (see AV2). This verse speaks of a positive message, and tells of a time when Adele’s character was happy (“yesterday was the time of our lives”). The fact that these memories are so vivid and fresh in the character’s memory, like they were “only yesterday”, only serves to heighten the sense of pain and sadness expressed elsewhere in the song (see AV2).

Here, the emotional message between lyrics and vocal timbre is much more aligned (see Figure 8.26). The lyrics convey a positive message. The vocal timbre too is much more resolved and controlled. This is evident through less variation being present through the vocal features of tense/lax, high/low, and loud/soft. It is also shown through the minimal use of creak and aspirate onsets (see Figure 8.22), which were used in verse one to indicate pain and sadness. In this way, the overarching emotional message in verse two is easily accessible (see Figure 8.26), allowing the listener to enjoy the positive emotions—as respite from the pain and sadness expressed in verse one and (as will be discussed in 8.3.2.3) the tension and grief expressed in the first chorus. This emotional message is likely somewhat arousing as many listeners would be able to associate with the sense of happiness (see Figure 8.27), however this arousal may be moderated by the fact that the vocal timbre used here remains quite neutral—Adele’s character does not completely give herself over to this emotion.

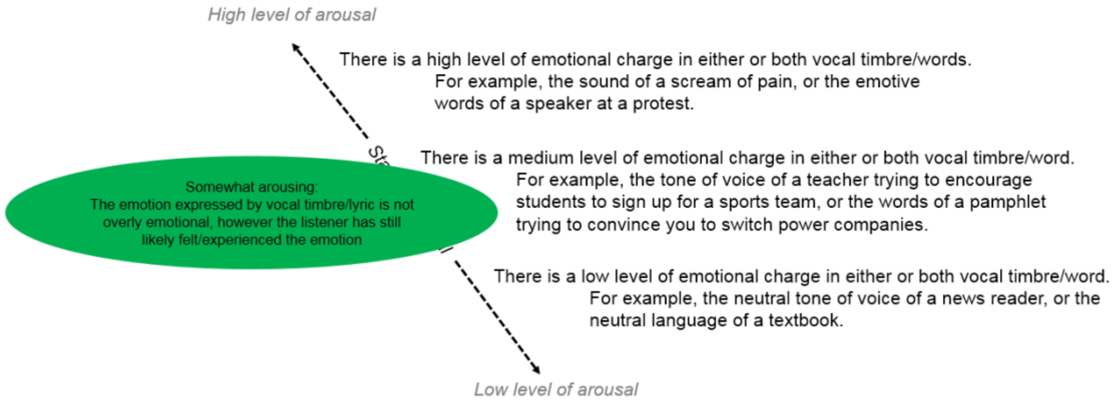

High level of arousal
There is a high level of emotional charge in either or both vocal timbre/words. For example, the sound of a scream of pain, or the emotive words of a speaker at a protest.
There is a medium level of emotional charge in either or both vocal timbre/word. For example, the tone of voice of a teacher trying to encourage students to sign up for a sports team, or the words of a pamphlet trying to convince you to switch power companies.
Somewhat arousing:
The emotion expressed by vocal timbre/lyric is not overly emotional, however the listener has still likely felt/experienced the emotion
There is a low level of emotional charge in either or both vocal timbre/word. For example, the neutral tone of voice of a news reader, or the neutral language of a textbook.
Low level of arousal

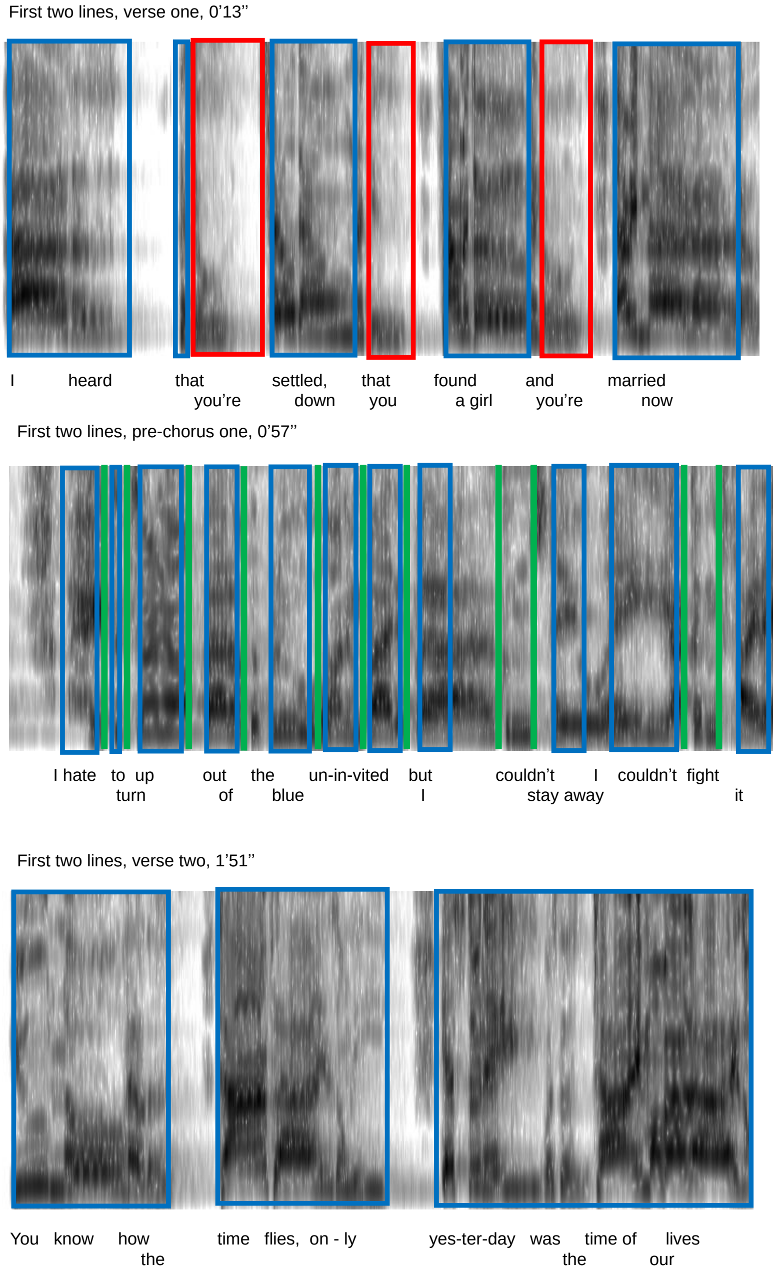
Figure 8.27. Attachment in “Someone” in verse two, and in the prechorus and Bridge. The highlighted sections show the likely level of arousal evoked by the emotive content of vocal timbre and lyrics.

8.3.2.2 Prechorus one and two, 0’57” – 1’12” and 2’19” – 2’36”, and Bridge 3’05” – 3’23”.

The first and second prechoruses, and the Bridge, are presented in much the same vocal timbre as heard in verse two. These sections are delivered in a manner that is generally smooth, plain, and clear, moderate in tension, volume, and register, and with mostly simultaneous onsets (see AV2, 0’57” – 1’12”, 2’19” – 2’36”). These musical features remain consistent throughout the prechoruses and Bridge, evoking a neutral, controlled emotional response.

The one vocal feature which contrasts with the timbre of verse two is termination. In these sections, terminations are very clear and crisp. While this has the effect of making the lyrics easily understandable (as each word is clearly separated), when coupled with the otherwise neutral vocal features it has the effect of making the speech sound artificial. The speech is too clear, too controlled, too well separated, for a vocal timbre that is otherwise relaxed.

Figure 8.28 shows a spectrograph of the first two lines of prechorus one, as compared with the first two lines of verses one and two. The constant use of stronger intensities in the prechorus can be seen. This is similar to the vocal timbre presented in the first two lines of verse two. However, the first two lines of the prechorus also show clear terminations, a feature not seen in either first or second verses. In this way, the vocal timbre in the prechorus (and also in the Bridge which shares vocal timbre features with this section) is very controlled and precise, in a way not heard in either of the verses.


First two lines, verse one, 0'13''

First two lines, pre-chorus one, 0'57''

First two lines, verse two, 1'51''

Words sung with stronger intensities are shown in blue boxes, words sung with weaker intensities are shown in red boxes.

Clear terminations are shown by the green line.

The first line (pre-chorus one) shows a consistent vocal timbre and clear terminations, compared to the second line (verse one)

Figure 8.28. Spectrographs of the first two lines of prechorus one, as compared to the first two lines of verse one and two. Differences in sustain and termination can be seen between the three spectrographs. The single quotation mark (') represents minute/s and the double quotation mark (") represents second/s.

The lyrics do not necessarily align with this message of hyper-controlled neutrality. Instead, the lyrics in the prechoruses explicitly speak of the inability of Adele’s character to control her emotions:

> I hate to turn up out of the blue uninvited
> But I couldn't stay away, I couldn't fight it.
> I'd hoped you'd see my face and that you'd be reminded
> That for me it isn't over. (Adkins & Wilson, 2011, track 11)

The first two lines speak of how she has turned up on impulse (she “couldn’t fight it”), unannounced and without reason, on her ex-boyfriend’s (and possibly his new wife’s) doorstep. The next two lines have a pleading tone. By re-inserting herself in this way, Adele’s character had hoped to remind her ex-boyfriend that, for her, the relationship “isn’t over”. However, what she hoped to achieve through such a reminder is unclear. Was it to rekindle the spark? Did she wish for him to see the pain he had caused? Did she wish to cause him pain?

The lyrics in the Bridge carry a similar message, however this time they are more rhetorical. We get the sense that in the prechorus Adele’s character is speaking directly to her ex-boyfriend, whereas in the Bridge she is speaking to herself:

> Nothing compares
> No worries or cares
> Regrets and mistakes

> They are memories made.
>
> Who would have known how bittersweet this would taste. (Adkins & Wilson, 2011, track 11)

These lyrics give the impression that Adele's character, while pained, is making an attempt to rationalise this experience and to resign herself to her pain and suffering. After all, in her words "regrets and mistakes" are "memories made". However, she is perhaps not as successful as we may suppose as the memory is still "bittersweet", still painful.

In general, across the prechoruses and Bridge, the impulsive, pleading tone of the lyrics is slightly misaligned with the meticulously controlled neutrality conveyed in the vocal timbre (see Figure 8.24). This creates a somewhat unsettling emotional message as the listener reconciles the sense of pain, pleading and resignation evoked through the lyrics with the hyper-controlled timbre. This creates a sense of detachment (one can become numb and detached when experiencing extreme sadness, this relationship is mapped in Figure 8.29). The emotional message is ambiguous—we are not sure how to interpret this sense of detachment created in these sections (see Figure 8.24). While confused, the emotive content of the lyrics nonetheless mean that the listener is still likely somewhat aroused by the overarching emotional message expressed in the prechorus and Bridge (see Figure 8.27).

8.3.2.3 Choruses.

Different emotions have been presented in the verses, prechoruses, and Bridge discussed so far in this chapter. In the chorus, we see not only the presentation of certain emotions, but the *evolution* of the character's emotional state. Adele's delivery of the chorus gradually betrays the character's increasing desperation (as illustrated through the words "don't forget me, I beg") and resignation (the lyrics "never mind, I'll find someone like you" suggest that she is resigned to her fate) (see AV2) (these themes are mapped in Figure 8.29). The lyrics of the choruses remain mostly the same, with only a few small changes:

Chorus 1 and 4

Never mind, I'll find someone like you.
I wish nothing but the best for you too.
Don't forget me, I beg.
I'll remember you said,
"Sometimes it lasts in love but sometimes it hurts instead.
Sometimes it lasts in love but sometimes it hurts instead." (Adkins & Wilson, 2011, track 11)

Chorus 2

Never mind, I'll find someone like you.
I wish nothing but the best for you too.
Don't forget me, I beg.

I'll remember you said,
"Sometimes it lasts in love but sometimes it hurts instead." (Adkins & Wilson, 2011, track 11)

Chorus 3

Never mind, I'll find someone like you.
I wish nothing but the best for you.
Don't forget me, I beg.
I'll remember you said,
"Sometimes it lasts in love but sometimes it hurts instead." (Adkins & Wilson, 2011, track 11)

To understand how Adele's character moves through a range of emotional themes in the chorus, the treatment of each individual line will be addressed here. Rather than reproducing individual spectrographs, the analysis of this chorus is best read in conjunction with AV2.

Never mind, I'll find someone like you (Adkins & Wilson, 2011, track 11)

In the first two choruses (beginning 1'13" and 2'37"), this line is delivered in a high, smooth, clear, and strong vocal timbre. Adele's voice is neither very lax nor very tense, and a moderate level of vibrato can be heard (see AV2). This, combined with the use of simultaneous onsets and clear terminations, gives the sense that Adele’s character is coping well with the end of the relationship. She seems in control of her emotions and we believe her when she tells us to “never mind” her, she’ll be fine, and she’ll soon find “someone like” her ex-boyfriend.

In the third chorus (3’24”), Adele’s character’s underlying emotion begins to surface. This is shown through a more varied vocal delivery. Adele sings this line, like in the first two choruses, high and strong. However, her levels of tension and breathiness show increasing variation, and the sound is rougher overall. It is as though the character is still trying to convince the listener, and herself, that she is “fine”, but, in the same way that one’s voice can break with emotion when speaking, her vocal timbre is beginning to break under the strain of maintaining the pretence. Sadness and desperation are creeping in.

In chorus four, Adele’s vocal timbre returns to the high, clear, strong sound heard in the opening two choruses (3’52”). She seems to have regained control of her emotions, however the sadness and desperation that appeared in chorus three remain there, just under the surface. This is shown in the overall rougher sound of Adele’s voice, and in her treatment of the words “someone” (see AV2). “Someone” retains the rough, and tense sound heard in chorus three. While we want to believe that the strong, high, and clear opening means that Adele’s character is “back on track”, these hints of tension and roughness suggest otherwise.

I wish nothing but the best for you (too) (Adkins & Wilson, 2011, track 11).

The delivery of this line follows a similar pattern to that discussed in the preceding paragraphs. The first two choruses are loud, clear, and strong, with a medium level of tenseness. The line teeters between plain and slight vibrato, and smooth and slightly rough. Chorus three, however, is more breathy and the level of tension is much more varied. It is as though Adele's character has become so sad, so melancholic (this relationship is included in Figure 8.29) that she no longer can, or no longer has the energy to, present a controlled and resolute front. While her voice is emotive, the lyrical message is still sincere. Like in the first two choruses, we do believe that she wishes him "nothing but the best", however the melancholic vocal sound used in this chorus gives this line a new under tone. This time she is saying "I wish you the best but remember, it's at my expense"—to make his "dreams come true" (i.e., to marry a girl who gives him things that Adele's character couldn't) she must have her heart broken, but she is prepared to martyr herself in this way for the sake of his happiness. This change in emotional message is underscored by the slight variation in lyrics from "I wish nothing but the best for *you too* [emphasis added]" to "I wish nothing but the best for *you* [emphasis added]". The first lyric implies that the feeling is mutual (he wishes her the best, and she wishes him the best too). The second eliminates any suggestion of wellwishing from the ex-boyfriend, and instead implies that it is only Adele's character who has made this gesture, heightening her sense of sacrifice.

In chorus four, there is a return to the stronger, louder, tenser sound. However, this line now features increasing levels of roughness. This is particularly evident on the words "best" and "for", which are especially tense and rough (see AV2). This increased tenseness and roughness, contrasted with what is otherwise a timbre quite like that heard in the first two choruses, once more creates the feeling that Adele's character is trying to convince the listener that she is emotionally stable, however her voice is again beginning to "crack".

> *Don't forget me, I beg. I'll remember you said, (Adkins & Wilson, 2011, track 11)*

This line shows a particularly noticeable increase in tension, roughness, and breathiness throughout the four choruses. In the first and second choruses, this line is overall tenser and breathier than that heard in "I wish nothing but the best for you" discussed in the previous paragraph. This juxtaposition serves to heighten the sense of desperation in Adele's character's voice as she pleads (begs) with him not to forget her. Contrast in vocal timbre can also be heard within this line. The "I" of "I beg", for example, stands out as it is very tense, rough, and breathy (that this emphasis is added to the word "I" is potentially significant for the emotional message and it will be discussed in more detail in the following paragraph). By contrasting these moments of very tense, rough, and breathy timbres with a vocal sound that is already quite tense and rough, not only does this section evoke a sense of grief, but it suggests that the character is now unable to control her emotions. To prioritise her linguistic message, and make sure that the listener understands how important it is to her that she not be forgotten, Adele's character has had to sacrifice some of the illusion of composure established in the beginning of these choruses.

When this line is presented in the third chorus, the sound is much rougher. Adele's character is not only grieving the end of a relationship, but she is desperate not to be forgotten. This desperation is heightened when we hear this line presented in the fourth chorus. Until this point, each lyric line has followed a general pattern of:

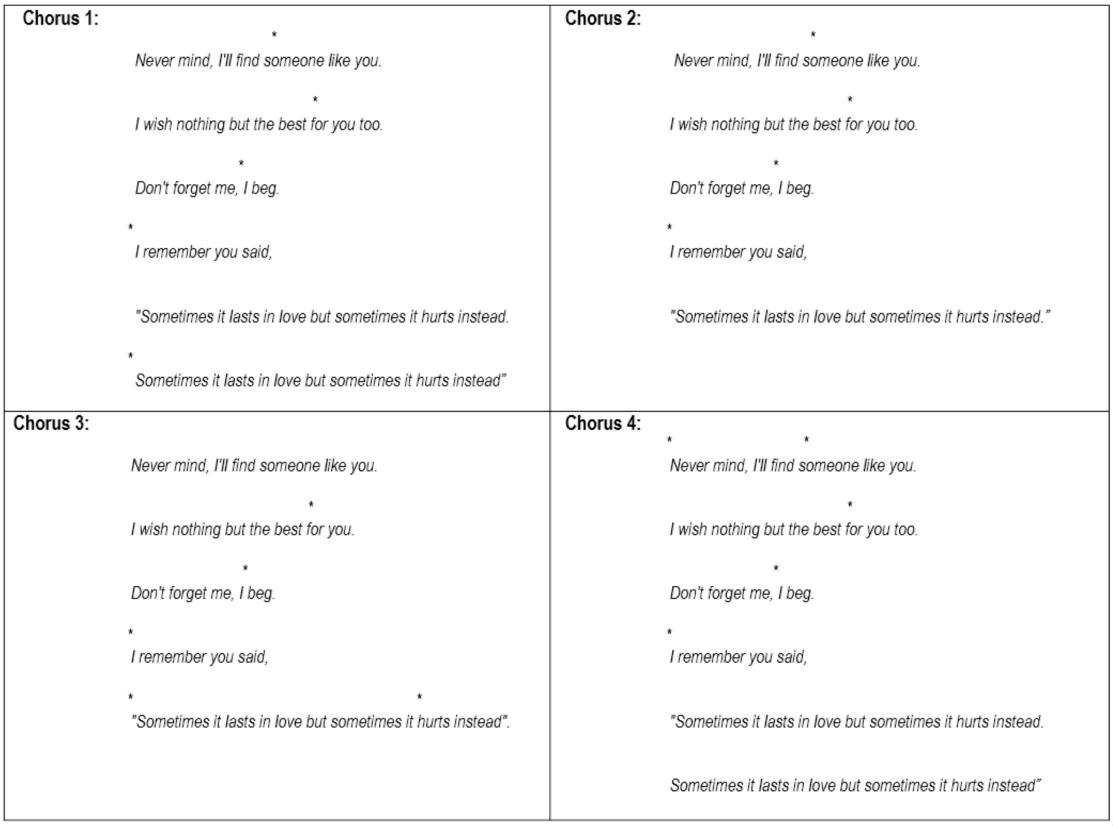
- Emotional but controlled timbre in chorus one and two,
- Emotion beginning to get out of hand in chorus three,
- Some control being regained in chorus four.

However, this line breaks that pattern. Control is not regained in this presentation of “don’t forget me I beg, I remember you said” (see AV2). Rather it sounds higher, rougher, and tenser, and a noticeable increase in audible breaths can be heard. One place in which this can be seen is, as touched on in section 8.3.2.1, is in the word “I”. In chorus four, Adele almost screams while singing “I” (see AV2). This treatment of “I”, beginning in the first two choruses and becoming increasingly louder and rougher in the third and fourth choruses, is grating. *I* is a word one uses to refer to oneself, and it may be imbued with connotations tied up in how that person sees themselves. In this way, *I* is not only a personal pronoun, but it may also be a linguistic representation of one’s idea of oneself. To hear a word which may be so intertwined with ideas of personal identity sung in such a noisy, grating, way, is unsettling. The vocal timbre with which this word is sung conflicts with the nuances of personal identity this word conveys, making it sound instead primal, basic, and exposed. And this is how Adele’s character sounds here—it is as though the ending of the relationship (supposedly by the ex-boyfriend as she is the one chasing after him in this song) has torn from her her own sense of personal identity, leaving her emotionally bare. In this way, her vocal timbre here may be an aural representation of her emotional state—Adele's character has not (could not?) regained control of her emotions, but rather her emotions are beginning to take control of her.

Breath may further emphasize this emotional message. The words "I beg" are bracketed by audible breaths (see AV2). The use of breaths before and after "I beg" is common to all four choruses (see Table 8.1). However, it is not necessary. Adele demonstrates throughout "Someone" that she is able to sustain long phrases on a single breath (e.g., see the breathing pattern for the lines "sometimes it lasts in love, sometimes it hurts instead" in choruses one, two, and four, Table 8.1). These breaths, then, are unexpected. They evoke negative (i.e., sad) emotional connotations such as a sense of grief (like an explosive and uncontrollable sob) and despair (like the raggedy breath heard in a panicked voice) (this connection is depicted in Figure 8.29). In this way, these breaths seem to corroborate the emotional message expressed by Adele's character in the words "I beg" grief in the first two choruses, desperation in the last two choruses.

Table 8.1.

Audible breath placement (indicated by asterisks) in the four choruses of "Someone" (Adkins & Wilson, 2011, track 11).



Chorus 1:

```
                       *
 Never mind, I'll find someone like you.
                             *
 I wish nothing but the best for you too.
                 *
 Don't forget me, I beg.
*
 I remember you said,

 "Sometimes it lasts in love but sometimes it hurts instead.
*
 Sometimes it lasts in love but sometimes it hurts instead"
```

Chorus 2:

```
                       *
 Never mind, I'll find someone like you.
                             *
 I wish nothing but the best for you too.
                 *
 Don't forget me, I beg.
*
 I remember you said,

 "Sometimes it lasts in love but sometimes it hurts instead."
```

Chorus 3:

```

 Never mind, I'll find someone like you.
                             *
 I wish nothing but the best for you.
                 *
 Don't forget me, I beg.
*
 I remember you said,
*                                            *
 "Sometimes it lasts in love but sometimes it hurts instead".
```

Chorus 4:

```
*                      *
 Never mind, I'll find someone like you.
                             *
 I wish nothing but the best for you too.
                 *
 Don't forget me, I beg.
*
 I remember you said,

 "Sometimes it lasts in love but sometimes it hurts instead.

 Sometimes it lasts in love but sometimes it hurts instead"
```

“Sometimes it lasts in love but sometimes it hurts instead

(Sometimes it lasts in love but sometimes it hurts instead)” (Adkins & Wilson, 2011, track 11)

These lines feature a different vocal timbre than that seen previously in the choruses. The tension and roughness, a feature of Adele’s vocal timbre in the line “Don't forget me, I beg. I'll remember you said”, is contrasted here with a smoother, and increasingly lax sound (see AV2). The difference in vocal timbre on this line compared to the others discussed above is interesting as, at this point, Adele’s character is not singing her own words. Rather she is singing *his* words—he was the one who told her, perhaps to ease the pain of the breakup, that “sometimes it lasts in love, sometimes it hurts instead”. The change in vocal timbre here may reflect this idea that Adele’s character is singing someone else’s words.

In the first chorus, this line begins relatively tense, strong, and high (1’33”). This vocal timbre continues into the second iteration of the line, with “sometimes it lasts in love” (1’41”) having a similar level of tension. By the lyrics “sometimes it hurts instead” (1’43”), however Adele’s vocal timbre becomes lax, low, weak, and breathy. A similar treatment of this line can be seen in chorus two. “Sometimes it lasts in love” at 2’57” is strong and tense, with “but sometimes it hurts instead” (3’00”) being sung in a softer, laxer, lower, and weaker timbre.

The presentation of this line in chorus three begins with a vocal timbre of moderate levels of tension and volume (3’44”). The timbre of the lyrics “sometimes it hurts instead” (3’47”), rather than becoming weaker and softer as in the previous choruses, continues with this use of tension, register, and volume. The delivery of this entire line in chorus three creates a sense of unwavering sadness. The line is not particularly energetic, nor is it especially weak and feeble. Rather, it conveys a kind of stagnant melancholy. The opening two choruses have created an expectation that there will be some change between the statement “sometimes it lasts in love” and the statement “sometimes it hurts instead” (see AV2). However, in chorus three the listener is not presented with this dichotomy, but rather with a vocal timbre that is generally unmoving (see 3’44” – 3’51”, AV2). This use of vocal timbre gives the impression that Adele’s character, already from the outset of this line, knows that she will not be able to convince herself of its truth.

In this chorus, Adele’s character sounds not only resigned, but also exasperated. This sense of exasperation is evoked by the placement of the breath just before the word “hurt” (see AV2, 3’48”, and Table 8.1, chorus three). Adele has not breathed in this place previously (and, as discussed above, likely does not need to breathe here). The presence of this unexpected breath gives the sense that Adele’s character is bracing herself against the tyranny of repeating the words which were thrown at her possibly by way of justification for the breakup. The breath also serves to emphasize the word “hurts” (see AV2). And indeed it sounds as though Adele’s character is suffering here, gasping for breath before singing “hurts” in a rough and breathy timbre (see AV2).

The presentation of this line in the fourth chorus (4’12”) combines ideas from the first three choruses. Here, the line is sung twice. The first presentation is like that of the first (first iteration of the line at 1’33”) and third choruses with the vocal timbre remaining consistent—in this case, consistently tense, high, and strong (see AV2, 4’12”–4’19”). The second presentation is more like that of the second iteration of the line in chorus one (1’41”) and the statement of the line in chorus two in that it gradually becomes more lax, breathy, rough, and weak (see AV2, 4’20”–4’30”).

The consistent vocal timbre, and even the increase in tension and pitch at 4’19”–4’20”, evoke a sense of resolve at the end of the first presentation of this line in chorus four. Perhaps Adele’s character has managed to accept the message these lyrics express—sometimes love hurts, it’s a fact of life. In the second presentation of this line, however, this resolution vanishes as quickly as it came. The breathy, and weak sustain, combined with the tapering termination, re-evoke the sense of grief and despair present earlier in the chorus. It is this grief and despair that the listener is left with at the end of the song, emotions evoked not by Adele’s character’s own story, but by the act of recalling someone else’s words.

General thoughts on the chorus

In general, the vocal timbre in the chorus explores a range of emotional themes and nuances. As the tension in Adele’s vocal timbre increases, feelings of desperation are evoked. These feelings are especially evident in the lines:

Don't forget me, I beg
I'll remember you said,
"Sometimes it lasts in love but sometimes it hurts instead,

> Sometimes it lasts in love but sometimes it hurts instead” (Adkins & Wilson, 2011, track 11)

The lyrics in these lines are also particularly emotive as they plead for the lost lover to not “forget me”, and they also repeat the words given to Adele’s character by her ex-boyfriend to, presumably, ease the pain of the breakup. However, these words are more of a consolation, rather than a justification. We don’t know why the couple broke up, and it is possible that Adele’s character doesn’t know either. This lack of closure may be the source of Adele’s grief and despair, as she can only guess at the reason for her heartbreak (as suggested by the line in verse one “I *guess* [emphasis added] she gave you things, I didn’t give to you”, which may translate to “I don’t know why I couldn’t make you happy, I don’t know what I did wrong”).

In this way, while all four choruses display an emotional progression from sad to desperate, they also present vocal timbre and lyrics that are aligned in emotional meaning, creating an easily accessible emotional message (see Figure 8.26). Therefore, the listener does not need to continually reassess the emotion present in vocal timbre and lyrics, but rather they can instead immerse themselves in the emotion created in the choruses—one of grief and despair. These emotions are likely highly arousing for the listener as relationship breakdowns are common human experiences (see Figure 8.25). Furthermore, most listeners could also identify with the frustration and despair caused by relationship breakdowns (romantic or otherwise) where we don't know the reason. In this way, the emotional message conveyed by vocal timbre and lyrics is easily accessible and highly arousing.

8.3.2.4 Concluding thoughts.

Emotional themes

Sadness is certainly a primary emotional theme in "Someone" (Adkins & Wilson, 2011, track 11). This theme has been identified based on both lyrics and vocal timbre. A number of nuances of this sadness have been identified in the above analysis. This is perhaps not surprising as sadness, especially when evoked by a relationship breakdown, is a multifaceted and complex emotion. The Emotional Map in Figure 8.29 provides a form of visual documentation of these emotional themes/nuances identified in "Someone".

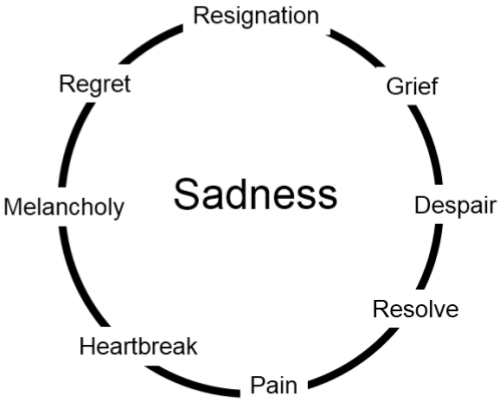


Figure 8.29. Emotional Map in "Someone" (Adkins & Wilson, 2011, track 11). This vocabulary set lays out the primary emotional themes and their nuances and variations.

Final thoughts

The above analysis has examined "Someone" by section, and has shown that each section presents distinct emotional messages. The verses are sad and unpredictable, but also feature control and resolution (in verse two). The prechorus and Bridge are unsettling, disconnected, and detached, but also feature Adele's character trying to make sense of her emotions (as in the Bridge). The choruses show Adele's character's progression from sadness to desperation, ending the song with a sense of burnt-out hopelessness. While these emotional messages are interesting in themselves, further insight can be gained into the overarching emotional message of "Someone" when these sections are viewed chronologically in the global context (i.e., in relation to one and other).

In this case, the following emotional pattern presents itself (this is visualised in Figure 8.30):

- Verse 1: giving the illusion of control, but unpredictably wavering from sadness to resolution.
- Prechorus 1: detached. Adele's character cannot come to terms with what has happened.
- Chorus 1: mostly in control, but a deep sense of sadness is evident.

- Verse 2: resolved, controlled, Adele’s character is singing about the past, a time when she was safe and happy.
- Prechorus 2: detached. Adele’s character cannot come to terms with what has happened.
- Chorus 2: mostly in control, but a deep sense of sadness is evident.
- Bridge: detached. Adele’s character is trying to come to terms with what has happened.
- Chorus 3: sadness is present, but it has manifested into desperation.
- Chorus 4: we expect Adele’s character to regain control here, but by the end of this chorus it is clear that the desperation and sadness are all that is left.

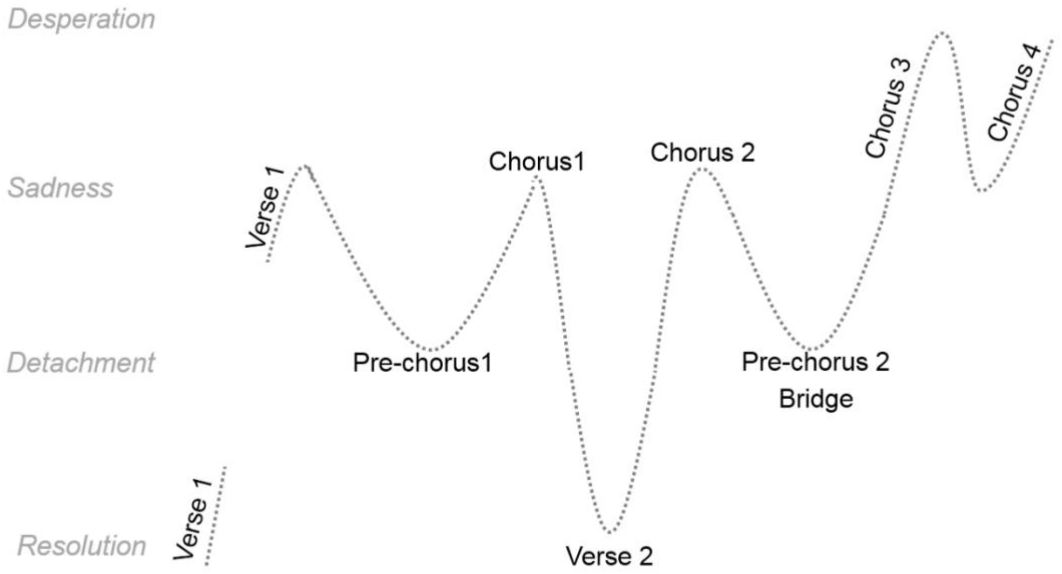


Figure 8.30. A visualisation of the pattern of emotional changes present throughout “Someone”.

In the global contexts, then, the emotional message between vocal timbre and lyrics is misaligned. Periods of clarity, calm, and neutrality (as in verse two, prechoruses, and Bridge) are contrasted with the increasingly frantic chorus. This juxtaposition foregrounds and heightens the feeling of unpredictability established in the first verse—this emotional instability hinted at in the local context is writ large in the global context.

Why is a sense of control evident in all but the choruses? One reason may be that the verses are directed at her ex-boyfriend. Therefore, in an attempt to not frighten him off, Adele's character tries to maintain the illusion of control. In the choruses, however, she is speaking to herself. Therefore, she can allow herself more emotional freedom. Whatever the reason, the emotional progression of this song is not linear, rather it is a tapestry. This seems appropriate as grief and heartbreak are not straightforward emotional experiences, but rather they are experiences in which the emotions seem to have uncontrollable, unpredictable, minds of their own.

8.4 Conclusion

This chapter shows how the analytical technique developed in this thesis may be applied in analyses. Though the methodology outlined in Chapter 7 can be seen to be appropriate for the purpose of analysing timbre as a contributor of emotional meaning in popular song, the application of the technique to real cases has revealed a few areas that require further work. One of these is exploring ways in which the analysis can be applied to more than one voice at a time. Another is the need to streamline the creation of the video descriptions through the use of computer automation.

While there is room for further development, this chapter has shown that, in its current form, this analytical technique provides the tools to conduct in-depth and meaningful analyses of vocal timbre based on how it conveys emotion, and how this impacts a listener's emotional perception of lyrics.

9 Conclusions and Further Research

9.1 Closing Thoughts

Vocal timbre is an evocative and salient sound source which is increasingly being included in academic study and musical analysis. The present thesis has contributed to this field in several ways. By showing that emotion conveyed by the vocal timbre reliably affected how listeners responded to sung words in a way that was consistent across a cohort of listeners, this thesis has validated the claim that timbre does influence a listener's perception of emotional valence in lyrics. In other words, it has shown that the experience of emotion in timbre, and the way this impacts emotional perception of lyrics, may be considered *intersubjective*. Therefore, analyses of vocal timbre which centre on its emotional content (as in this thesis) may be more than just a reflection of the analyser's experience of timbre (i.e., it may go beyond their idiosyncratic experience).

This thesis has also offered evidence that the form of certain acoustic features within a vocal timbre (i.e., the onset, the sustain, and the termination of a note) are good indicators of particular emotional valences. Such findings suggest that, within an emotionally valenced vocal timbre, there are aural cues that can act as predictors of emotional valence in a robust way. The identification of such cues should provide a way of classifying emotional valence in vocal timbre, something which is important as, presently, no such system exists. In the identification of specific emotional acoustic features of a vocal timbre, this thesis has adopted an approach based on the research presented here as well as that of van Leeuwen (1999), Heidemann (2016) and Jo Estill (see McDonald Kilmek et al., 2005). I have called these acoustic features the Vocal Timbre Features. Although they require further refinement, the Vocal Timbre Features have shown value in helping to reliably and robustly identify emotional valence in vocal timbre.

One reason put forward for vocal timbre having the potential to effectively convey emotion, and for each aspect of the Vocal Timbre Features to be considered characteristic of certain emotional states, is the presence of noise and other nonpitched sounds within a vocal timbre. It has been argued that the more noisy, nonpitched sounds present within a vocal timbre, the more likely it is to be perceived as conveying a negative emotion. This is likely to be linked to our everyday lives where noise tends to be linked with negative experiences. In this way, this thesis has also suggested that the way we interpret emotional meaning in timbre may be rooted in, and shaped by, our real life experiences of sound. All these findings support the idea that vocal timbre is a salient and meaningful musical feature. While this may be true for all music, it is especially true for popular vocal songs which rely heavily on the vocal line for musical expression.

These findings culminated in the development and implementation of a new analytical technique for vocal timbre. This analytical technique allows for vocal timbre to be described and annotated in audiovisual format. Such a format has been especially useful for vocal timbre analysis as it allows one to see the vocal timbre being described (through spectrographs and Vocal Timbre Features) while at the same time hearing the vocal timbre in question. The technique also includes several diagrammatic vocabulary sets which allow one to describe the synergies and document the relationship between emotions conveyed through sung lyrics and vocal timbre, and to make predictions about the level of arousal they may evoke in a listener.

The technique developed here is particularly useful for recording-based analyses as it is multilayered. That is, it draws on a number of tools and texts (from the recording itself, to the annotation of vocal timbre through the Vocal Timbre Features) to facilitate analysis. This multilayered approach could be useful for the analysis of other musical features which are most often or most accurately preserved through the recording.

9.2 An understanding of vocal timbre – revisited

1. In section 2.2 a working definition of vocal timbre was outlined. Given the new understanding of vocal timbre arrived at in this thesis, it is timely now to revisit some of the concepts that have been found to be important for discussing and analysing vocal timbre. These are: Vocal timbre is the characteristic sound of a singer's voice that differentiates it from other sound sources (and indeed from the sound of other singers), even when they share the same volume (dynamic), frequency (pitch), and time (duration);

2. Vocal timbre consists of a variety of pitched and unpitched vocal sounds (breathiness, stylised screams/cries, throaty sounds, etc.) derived from the way in which a single singer may use their body to produce sound for musical expression
3. Vocal timbre has the potential to evoke an emotional response, and to evoke a connection between the emotion expressed in the sung voice and a real life experience (memory) such that source bonding (hearing a vocal timbre and relating it to a life experience, either consciously or subconsciously) impacts the perception of meaning; and,
4. The production and sound of a vocal timbre can be impacted by other musical features, which (although not timbral themselves) should be considered in the analysis of vocal timbre.

Two concepts have been added in this revised definition. First, that vocal timbre has the potential to evoke an emotional response through eliciting memories of a real life experience, and second that some aspects which might be considered extra-timbral may be important when discussing vocal timbre.

Timbre and source bonding

This revised definition suggests that vocal timbre may be emotionally evocative because of the possible connection between a sung vocal utterance and our “real life” experiences of sound. The concept of *source bonding* offers one way to explain this idea. Source bonding is a term coined by Denis Smalley that refers to:

> the natural tendency to relate sounds to supposed sources and causes, and to relate sounds to each other because they appear to have shared or associated origins. (Smalley, 1994, p. 37)

This concept may be adapted to vocal timbre research and used to refer to aspects of a sung voice that evoke memories/emotional responses from other, real life, situations (either consciously or subconsciously). One way in which this may be understood is by looking at para-linguistic features, which play a key role in conveying emotion in spoken language (see, for example, Poyatos, 1993, 2002). A connection between singing and speech has already been drawn in earlier research (see, for example, Lacasse, 2010b, and Middleton, 2000). Therefore, source bonding is likely to play a role in a listener’s experience of a vocal timbre in the same way that para-linguistic features of speech do when decoding vocal timbre.

Timbre and other musical features

Timbre is a “multidimensional attribute” (Plomp, 1970, as cited in Rossing, 1990, p. 126). The discussion of timbre in this thesis has highlighted the complex and dynamic nature of this musical feature. The definition of timbre offered in section 2.2 examined timbre in two ways: as a technical, psychoacoustic phenomenon, and as a phenomenological one. In that discussion, it was recognised that other musical features play a role in the production of timbre (e.g., frequency, as discussed in Rossing, 1990, pp. 30 - 31). This idea was discussed in more detail in section 7.3 where the impact of frequency and dynamic on vocal timbre was explored. These musical features, although not timbral themselves, impact on a vocal timbre's sound. Therefore, the need to consider such extra-timbral elements in study of timbre has been recognised in this thesis.

As with any large-scale study, many avenues for future research present themselves. These avenues could shed further light on the nature of emotional perception in vocal timbre, and help towards the development of new analytical, more inclusive, techniques (e.g., analytical techniques that can account for the role of vocal timbre in music other than popular vocal songs, or in instances where lyrics are not easily discernible). Some of the most promising areas of research are listed in the following section.

9.3 Avenues for Further Research

9.3.1 Refining a system for classifying emotional valence in vocal timbre.

The analytical technique developed in this thesis relies on assessing the emotion in both vocal timbre and lyrics. Much research has been done on the perceived emotional valence of words, and this can be used for classifying emotional valence in lyrics. But far less research has been done on perceived emotional valence in vocal timbre. While the development of the Vocal Timbre Features offers a way forward in terms of classifying emotional valence in vocal timbre, the system could be further refined to increase its usability and reliability.

Further developing the Vocal Timbre Features.

In Part II of this thesis, certain Timbral Attributes (namely, onset, sustain, and termination) were shown to be good predictors of how participants would perceive emotional valence in vocal timbre. One aspect of these attributes which seemed to influence how sad/negative a vocal timbre was perceived to be was noise. In Chapter 7, these Timbral Attributes, and the idea that noise can contribute to perception of negative emotions, were expanded on in the development of the Vocal Timbre Features. These Features have been useful for describing and identifying valence in vocal timbre. However, as the reception tests conducted in this thesis were not specifically designed to test the reliability of noise as a signal for negative emotions, or the reliability of the Timbral Attributes (or subsequent Vocal Timbre Features) as predictors, more research is needed to examine the precise nature of these emotional signals.

Exploring learned associations and embodiment.

It has been suggested that para-linguistic, nonlinguistic, and musical cues may transmit meaning and emotion because we understand others as we understand ourselves. In other words, to understand a sound we do not only listen to it, we internalise it, and, in so doing, we replicate the (emotional) conditions under which the sound was produced. There are two areas of research which explore this idea and which may yield results for vocal timbre too.

Learned associations.

Learned associations are related to the concept of source bonding. Simply put, it is the idea that we have, through previous experiences, learned that certain sounds (e.g., laughter) are associated with certain emotions and meanings (e.g., happiness). If these learned associations become deeply entrenched in our day to day emotional experiences, they may then be carried over into musical experiences too.

These learned associations also have the potential to be similar across listeners. Already in the 18th century philosopher Edmund Burke, wrote that:

> All the natural powers in man, which I know, that are conversant about external objects, are the senses; the imagination; and the judgment. And first with regard to the senses. We do and we must suppose, that as the conformation of their organs are nearly or altogether the same in all men, so the manner of perceiving external objects is in all men the same, or with little difference. (Burke, 1775/1824, p. 6)

To evidence these claims, Burke states that if every person were to experience the senses in a different way, then a consensus of thinking could never be reached. That is, our perceptions of objects and things is (for the most part) similar because our senses give us all the same information. For example, sugar is sweet, lemons are sour. If we did not have these same perceptions, Burke argues, we would never form consensus on any topic.

The same basic idea still exists today. For example, a similar argument can be found in Lakoff and Johnsen (2003). The authors argue that the widespread use of metaphor not only points to similar experiences of the senses, but also to the importance of the experience of the body and the environment on understanding and perception (Lakoff & Johnsen, 2003). For example, consider the link between the bodily experience and the perception of the mind in the expression "I'm *warming* up to her". A number of such body-mind connections have been found—e.g., a link between perceived movement through space and perceived movement through time (e.g., *future* = *forward*) (Miles, Stuart, & Macrae, 2011). Connections have also been found between the physical experience of tough and soft, and perceiving a face as being male and female (Slepian, Weisbuch, Rule, & Ambady, 2011). Such studies suggest that a connection between bodily experience and abstract cognition exist, and that it may be deeply rooted in how we understand the word. If this connection is robust and universal, then it is not implausible that it would carry into our musical experiences too.

Embodiment

Traditionally, cognitive science has taken the view that cognition is abstract. That is, the workings of the mind are separate to the functions of the body. More recently, however, cognitive research has started considering how the functions of the body, and its interactions with environment, impact cognition. This field of study is called embodied cognition. Embodied cognition “grants the body a central role in shaping the mind” (Wilson, 2002, p. 625). If it is that the interpretation of emotions in abstract cases (such as music) depends on their physical manifestations, then it is possible to see how features of a vocal timbre that are intrinsically related to particular emotions will come to play a part in how meaning is understood at a more conceptual level.

The idea that we may, either consciously or nonconsciously, interpret the emotions of others by replicating those emotions in ourselves is supported by the function of the mirror neuron system. Mirror neurons are “so named because they ‘mirror’ actions of other individuals by re-enacting them on the observer’s motor repertoire” (K. Keysers & Fadiga, 2008, p. 193). That is, the same neurons activated when doing a task are activated when observing the task. This phenomenon has been observed in humans in relation to the visual (Cattaneo & Rizzolatti, 2009; Proverbio, Riva, & Zani, 2009) and auditory (C. Keysers et al., 2003; Kohler et al., 2002) domains. The more experience one has with an action/sound, the stronger the mirror neutron response (Calvo-Merino, Glaser, Grezes, Passingham, & Haggard, 2005). It has also been suggested that “aspects of musical experience may be mediated by the human mirror neuron system” (Molnar-Szakacs & Overy, 2006, p. 235). In other words, music may be processed through the mirror neuron system in such a way that its perception becomes experiential rather than representational (Molnar-Szakacs & Overy, 2006, p. 239).

When considering vocal timbre in light of such research, it seems possible that a listener perceives meaning and emotion in vocal timbre through such processes. Vocal timbre is created by the voice. It is of the body. To change one's vocal timbre, one needs to change one's body (e.g., change the facial expression or manipulate the vocal tract to produce a specific vocal timbre). These bodily experiences and cognitive representations may be so intertwined that they are enacted in both day-to-day and in more abstract experiences.

By continuing to refine the ideas presented in this thesis, we will gain a better understanding of the role of vocal timbre in song. Such an understanding would help develop a framework whereby emotional valence in vocal timbre can be defined as quickly and efficiently as a harmonic progression can be identified.

9.3.2 The impact of non-words and multiple vocals on perception.

In the case studies presented in Chapter 8, two situations presented themselves which should be investigated in future research. The first is the presence of non-words. The research conducted in this thesis has been focused on emotional vocal timbres and emotional words. However, non-words do occur in song. Therefore, to extend the analytical technique for vocal timbre developed here, further research should examine the potential impact of timbre and non-words on emotion perception.

The second situation encountered is the presence of multiple voices at once (e.g., vocal harmonies). Again, research conducted in this thesis focused on solo voice. However, vocal harmonies are present in much music, and they often occur with distinguishable, and potentially important, lyrical content. Therefore, to expand the reach of this analytical technique, the impact of simultaneous voices on emotional perception of lyrics and vocal timbre should be investigated.

9.3.3 Vocal timbre in its wider musical environment.

As touched upon in section 8.2, the audio visual annotations of vocal timbre developed here are presently best suited to representing a single voice. However, multiple voices are often present in a song, and one may wish to annotate more than one voice in these instances. Further research could examine tools such as iAnalyse or EAnalylsis, which may facilitate the annotation of simultaneous vocal timbres in a song.

In the present study, I looked at vocal timbre and lyrics in relative isolation. However, vocal timbre almost always occurs alongside other musical features, such as instrumental accompaniment for example. Certainly, examining vocal timbre and lyrics in isolation is an important first step to better understanding how these musical features create and express emotion. However, a logical next step would be to examine if and how this emotion is mediated by other musical features within a song.

9.3.4 Listening environment.

The listening environment[24] presents an interesting challenge for recording-based analyses. If and how the listening environment impacts musical perception is not clear. However, because of the potential for listening environments to vary so substantially in both everyday listening contexts, and in analytical contexts, this is an area which would be worthy of further research.

9.3.5 Replicating analyses using the Vocal Timbre Features and automating the audio-visual annotations.

The Vocal Timbre Features outlined in this thesis are useful tools for analysing vocal timbre. However, there are two important areas in which further developing these features would be beneficial. First, in the present

[24] That is, the space in which we listen to music, such as over loudspeakers at a party, through headphones, in the car, and so on.

thesis, the Vocal Timbre Features were applied by myself only. In future, it would be beneficial to determine whether analyses conduced with these features are replicable. That is, to explore whether different analyses conducted by different analysers produces similar results. If different analyses conducted using the Vocal Timbre Features produce similar results, this will speak to their robustness as analytical tools.

Second, if the Vocal Timbre Features are consistently used by human analysers in a similar way, then it may be possible to develop a computer programme which produces similar results. That is, to automate the audio-visual annotations used in chapter 8. This would make such analyses much more efficient (as conducting the audio-visual annotations by hand is a very time consuming process), and allow the analyser to search recordings to, for example, identify the number of glottal onsets in a song and when they occur.

9.4 In Summary

Interdisciplinary approaches to the study of music are becoming increasingly common, generating new ways of thinking about, analysing, and studying music. These interdisciplinary approaches are significant for vocal timbre analysis as they are facilitating the examination of vocal features which were previously too intangible for systematic study. The present thesis has taken advantage of this rising interdisciplinary trend in the development of a new analytical approach for vocal timbre. In doing so, it has exposed the complexities of studying vocal timbre and emotion. In particular, this thesis has been concerned with questions such as:

- Do listeners experience and perceive emotion in vocal timbre in an intersubjective way?
- Does this experience of vocal timbre impact emotional perception of lyrics?
- What acoustic cues may serve as emotional indicators in a vocal timbre?
- How can we identify and use these cues to efficiently and reliably identify emotional valence in vocal timbre?
- How can we document and discuss the relationship between emotional messages conveyed in vocal timbres and emotional messages conveyed in lyrics?

While the answers to some of these questions must be refined in future research, this thesis has presented a starting point for studying and analysing vocal timbre in terms of its emotive content. It has also gone some way to set the scene for subsequent investigations into emotional perception of vocal timbre—surely, a rich and diverse avenue of research with potential benefits for a number of disciplines, not least of which is musicology.

Appendix

List of Happy, Neutral, and Sad Words Used for Stimuli in Main Test

Happy		Sad		Neutral	
Norm	*Word*	*Norm*	*Word*	*Norm*	*Word*
8.72	paradise	1.61	sad	4.15	fur
8.21	happy	1.98	poison	4.16	lump
8.17	lucky	2.00	upset	4.17	dirt
8.10	cheer	2.13	pain	4.35	shadow
8.03	proud	2.25	loser	4.36	corner
7.86	party	2.28	regretful	4.39	plain
7.82	beauty	2.29	violent	4.48	alley
7.80	enjoyment	2.34	defeated	4.51	bus
7.80	triumph	2.39	starving	4.52	obey
8.56	humor	2.41	alone	4.56	stool
7.74	admired	2.43	despairing	4.58	errand
8.45	laughter	1.61	death	4.61	bench
8.38	win	2.47	filth	4.64	knot
8.37	comedy	1.69	grief	4.67	cliff
8.37	cash	1.70	failure	4.74	square
8.37	fun	1.88	gloom	4.74	rough
8.60	joy	1.90	hurt	4.75	glass
8.26	kiss	1.93	misery	4.77	reptile
				4.82	board
				4.83	curtains

References

Adkins, A. & Wilson, D. (2011). Someone like you [Recorded by Adele]. On *21* [CD, digital download]. London, UK: XL.

Adolphs, R. (2002). Neural systems for recognizing emotion. *Current Opinion in Neurobiology, 12*(2), 169–177. doi:http://dx.doi.org/10.1016/S0959-4388(02)00301-X

Adorno, T. W. (1985). On the fetish-character in music and the regression of listening. In A. Arato & E. Gebhardt (Eds.), *The Essential Frankfurt School Reader* (pp. 270–299). New York, USA: The Continuum Publishing Company.

The AHRC Research Centre for the History and Analysis of Recorded Music. (2009). Retrieved from http://www.charm.rhul.ac.uk/about/about.html

American Physical Society. (2011). January 10, 1919: Death of Wallace Sabine, pioneer of architectural acoustics. *American Physical Society NEWS, 20*(1), p. 2.

Anderson, S. A., & Fuller, G. B. (2010). Effect of music on reading comprehension of junior high school students. *School Psychology Quarterly, 25*(3), 178–187. doi:10.1037/a0021213

Aucouturiera, J., Johanssonb, P., Hallb, L., Segninid, R., Mercadiéf, L., & Watanabeg, K. (2015). Covert digital manipulation of vocal emotion alter speakers' emotional states in a congruent direction. *Proceedings from the National Academy of Sciences, 113*(4), 948–953. doi:10.1073/pnas.1506552113

Audacity Team. (1999–2015). Audacity® (Version 2.1.2) [Computer software]. Retrieved from https://www.audacityteam.org

Auslander, P. (2004). Performance analysis and popular Music: A manifesto. *Theatre Review, 14*(1), 1–13. doi:10.1080/1026716032000128674

Ball, P. (2007). Why the Greeks could hear plays from the back row: An ancient theatre filters out low-frequency background noise. Retrieved from https://www.nature.com/news/2007/070319/full/news070319-16.html

Bauer, W. (2007). Louis Armstrong's "Skid Dat De Dat": Timbral Organization in an Early Scat Solo. Jazz Perspectives, 1(2), 133-165. doi:10.1080/17494060701611809

Baumeister, R., Bratslavsky, E., Finkenauer, C., & Vohs, K. (2001). Bad is stronger than good. *Review of General Psychology, 5*(4), 323–370. doi:10.1037//1089-2680.5.4.323

Behne, K. (1997). The development of "Musikerleben" in adolescence: How and why young people listen to music. In I. Deliege & J. A. Sloboda (Eds.), *Perception and cognition of music* (pp. 134–151). East Sussex, UK: Psychology Press Ltd.

Bent, I. (1987). *Analysis.* London, UK: Macmillan.

Bigand, E., Tillmann, B., Poulin-Charronnat, B., & Manderlier, D. (2005). Repetition priming: Is music special? *The Quarterly Journal of Experimental Psychology, 58A*(8), 1347–1375. doi:10.1080/02724980443000601

Blackburn, M. (2009, June). *Composing from spectromorphological vocabulary: proposed application, pedagogy and metadata*. Paper presented at the Electronic Music Studies Conference 09, (EMS), Buenos Aires, Argentina. Retrieved from http://www.ems-network.org/ems09/papers/blackburn.pdf

Blacking, J. (1981). Making artistic popular music: The goal of true folk. In R. Middleton & D. Horn (Eds.), *Popular music 1: Folk or popular? Distinctions, influences, continuities* (pp. 9–14). Cambridge, UK: Press Syndicate.

Blake, D. K. (2012). Timbre as differentiation in indi music. *Journal of the Society for Music Theory, 18*(2). Retrieved from http://mtosmt.org/issues/mto.12.18.2/mto.12.18.2.blake.html#Beginning

Blake, E. C., & Cross, I. (2015). The acoustic and auditory contexts of human behavior. *Current Anthropology, 56*(1), 81–103.

Bonada, J., & Loscos, A. (2003, August). *Sample-based singing voice synthesizer by spectral concatenation.* Paper presented at the Proceedings of the Stockholm Music Acoustics Conference, Stockholm, Sweden. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.14.6410&rep=rep1&type=pdf

Bowen, M. (1970). Musical development in pop. In T. Cash (Ed.), *Anatomy of pop* (pp. 32–57). London, UK: British Broadcasting Company.

Brackett, D. (1995). *Interpreting popular music.* Cambridge, UK: Cambridge University Press.

Bradley, M. M., & Lang, P. J. (1999). *Affective norms for English words (ANEW): Instruction manual and affective ratings* (Technical Report C-1). The Center for Research in Psychophysiology. Gainesville, USA: University of Florida.

Bradley, M. M., & Lang, P. J. (2007). *The international affective digitized sounds (2nd ed.; IADS-2): Affective ratings of sounds and instruction manual.* (Technical report B-3). Gainesville, USA: University of Florida.

Breen, M. (1987). Fundamentalist music: The popular impulse. In M. Breen (Ed.), *Missing in action: Australian popular music in perspective* (pp. 8–31). Melbourne, Australia: Sybylla Womens Cooperative.

Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound.* Cambridge, USA: The MIT Press.

Brown, A., & Bassett, J. (2015, December). *Spatial responsiveness—Allowing an acoustic environment to direct spontaneous musical composition*. Paper presented at the 2nd Conference of the Australian Music & Psychology Society (AMPS), Sydney, Australia.

Burke, E. (1824). *A Philosophical Inquiry into the origin of our ideas of The Sublime and Beautiful: With an introductory discourse concerning Taste: and several other additions.* Retrieved from https://play.google.com/books/reader?id=RTEzjpDl_TIC&printsec=frontcover&output=reader&hl=en&pg=GBS.PP3 (1775)

Burling, R. (2000). Comprehension, production and conventionalisation in the origins of language. In C. Knight, M. Studdert-Kennedy, & J. R. Hurford (Eds.), *The evolutionary emergence of language: Social functions and the origins of linguistic form* (pp. 27–39). Cambridge, UK: Cambridge University Press.

Burnham, D., Kitamura, C., & Vollmer-Conna, U. (2002). What's new, Pussycat? On talking to babies and animals. *Science, 269*(5572). doi: 10.1126/science.1069587

Burns, L. (1997). Joanie' Get Angry: k.d. lang's Feminist Revision. In J. Covach & G. M. Boone (Eds.), Understanding Rock: Essays in Musical Analysis (pp. 93 - 112). New York, USA: Oxford University Press.

Burns, L. (2000). Analytical methodologies for rock music: Harmonic and voice leading strategies in Tori Arnos's "Crucify". In W. Everett (Ed.), *Expression in pop-rock music: A collection of critical and analytical essays* (pp. 63–92). New York, USA and London, UK: Routledge, Taylor and Francis Group.

Butler, J. (2010). Revolution [Recorded by The John Butler Trio]. On *April uprising* [CD]. Fremantle, Australia: Family Music Pty Ltd.

Carter, T. (2001, January 01). Word-painting. Grove Music Online. Ed. Retrieved 24 Aug. 2018, from http:////www.oxfordmusiconline.com/grovemusic/view/10.1093/gmo/9781561592630.001.0001/omo-9781561592630-e-0000030568.

Calvo-Merino, B., Glaser, D. E., Grezes, J., Passingham, R. E., & Haggard, P. (2005). Action observation and acquired motor skills: an FMRI study with expert dancers. *Cerebral Cortex, 15*(8). doi:10.1093/cercor/bhi007

Cattaneo, L., & Rizzolatti, G. (2009). The mirror neuron system. *Archives of Neurology, 66*(5), 557–560. doi:10.1001/archneurol.2009.41

Chate, S. (n.d.). *Jo Estill—The Estill voice model.* Retrieved from Thesingingvoice.com website: http://thesingingvoice.com/about/vocal-technique/jo-estill

Chion, M. (2012). The three listening modes. In J. Sterne (Ed.), *The sound studies reader*. Oxon, UK: Routledge.

Clarke, D., & Clarke, E. (Eds.). (2011). *Music and consciousness: Philosophical, psychological, and cultural perspectives*. Oxford, UK: Oxford University Press.

Clarkson, M. G., Clifton, R. K., & Perris, E. E. (1988). Infant timbre perception: Discrimination of spectral envelopes. *Perception & Psychophysics, 43*(1), 15–20.

Cogan, R. (1987). *New images of musical sound.* Cambridge, USA: Harvard University Press.

Cogan, R., & Escot, P. (1976). *Sonic design*. New Jersey, USA: Prentice Hall, Inc.

Cole, P. (1970). Lyrics in pop. In T. Cash (Ed.), *Anatomy of pop*. London, UK: British Broadcasting Company.

Collins, B. S., & Mees, I. M. (2013). *Practical phonetics and phonology: A resource book for students*. Abingdon, United Kingdom: Routledge

Cooke, D. (1982). *Vindications: Essays on romantic music*. Cambridge, UK: Cambridge University Press.

Cottrell, S. (2010). The rise and rise of phonomusicology. In A. Bayley (Ed.), *Recorded music: Performance, culture and technology*. New York, USA: Cambridge University Press.

Cox, A. (2011). Embodying music: Principles of the mimetic hypothesis. *Music Theory Online, 17*(2). Retrieved from http://www.mtosmt.org/issues/mto.11.17.2/mto.11.17.2.cox.html

De Backer, W. (2011). Somebody that I used to know [Recorded by Gotye, featuring Kimbra]. On *Making mirrors* [CD single, digital download, 7" vinyl (promotional only]). Merricks, Australia: Eleven.

Declercq, N. F., & Dekeyser, C. S. A. (2007). Acoustic diffraction effects at the Hellenistic amphitheater of Epidaurus: Seat rows responsible for the marvelous acoustics. *Journal of the Acoustical Society of America, 121*(4), 2011-2022.

Dolar, M. (2006). *A voice and nothing more.* Cambridge, USA: The MIT Press.

Dudescolded. (2010, December 24). *I'll Cover you (HD)* [Video file]. Retrieved from https://www.youtube.com/watch?v=CUY_st9c-QA

Douven, I. (2018). A Bayesian perspective on Likert scales and central tendency. Psychonomic Bulletin & Review, 25(3), 1203–1211.

Dylan, B. (1963). Blowin' in the wind [Recorded by Bob Dylan]. On *DYLAN* [CD]. New York, USA: Columbia/Legacy. (2007)

Ekkekakis, P. (2012) Affect, Mood, and Emotion. In T. Tenenbaum, R.C. Eklund and A. Kamata (Eds.), Measurement in Sport and Exercise Psychology (pp. 321–332). USA: Human Kinetics.

Ekman, P. (1999) Basic Emotions. In: T. Dalgleish and M. Power (Eds.), *Handbook of Cognition and Emotion* (pp. 45–60). Sussex, UK: John Wiley & Sons Ltd,.

Ekman, P. (1994). Strong evidence for universals in facial expressions: A reply to Russel's mistaken critique. *Psychological Bulletin, 115*(2), 268-287.

Ekman, P. (1992). Are There Basic Emotions? *Psychological Review, 99(3),* 550-553.

Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology, 17*(2), 124-129.

Erickson, R. (1975). *Sound structure in music.* Berkeley, Los Angeles, USA; London, UK: University of California Press.

Fernald, A. (1985). Four-month-old infants prefer to listen to motherese. *Infant Behavior and Development, 8*(2), 181–195.

Fernald, A., & Kuhl, P. (1987). Acoustic determinants of infant preference for motherese speech. *Infant Behavior and Development, 10*(3), 279–293.

Forte, A. (1995). *The American popular ballad of the Golden Era, 1924–1950.* New Jersey, USA: Princeton University Press.

Foster, J., & Foster, K. (2015). *DMDX (5.1.3.4)* [Computer Software]. Retrieved from http://www.u.arizona.edu/~jforster/dmdx/

Friendofpoodles. (2008, January 5). *The Windy City from Calamity Jane (1953)* [Video file]. Retrieved from https://www.youtube.com/watch?v=5MnUrhptPSo&list=RD5MnUrhptPSo

Frith, S. (1998). *Performing rites: On the value of popular music.* Cambridge,USA: Harvard University Press.

Fromkin, V., Rodman, R., & Hyams, N. (2011). *An introduction to language* (9th ed.).Boston, USA: Wadsworth.

Furnham, A. (1986). Response bias, social desirability and dissimulation. *Personality and Individual Differences, 7*(3), doi:10.1016/0191-8869(86)90014-0

Geringer, J. M., MacLeod, R. B., Madsen, C. K., & Napoles, J. (2015). Perception of melodic intonation in performances with and without vibrato. *Psychology of Music, 43*(5), 675–685.

Godøy, R. I. (2011). Sound-action awareness in music. In D. Clarke & E. Clarke (Eds.), *Music and consciousness: Philosophical, psychological, and cultural perspectives* (pp. 231—244). Oxford, UK: Oxford University Press

Gordon, I. (1951). Unforgettable [Recorded by Nat King Cole]. On *Unforgettable* [vinyl record]. California, USA: Capitol. (1954)

Gritten, A. (2013). [*Review of the book Music and consciousness: Philosophical, psychological, and cultural perspectives,* by D. Clarke & E. Clarke (Eds.)]. *Psychology of Music, 41*(4), 519–522.

Gross, J.J., (2010). The Future's So Bright, I Gotta Wear Shades. *Emotion Review, 2(3),* pp. 212–216.

Grossmann, T., Oberecker, R., Koch, S. P., & Friederici, A. D. (2010). The developmental origins of voice processing in the human brain. *Neuron, 65*(6), 852–858.

Harran, D. (1986). Word-Tone Relations in Musical Thought from Antiquity to the Seventeenth Century. Rome: Hanssler-Verlag.

Hajda, J. M., Kendall, R. A., Carterette, E. C., & Harshberger, M. L. (1997). Methodological issue in timbre research. In I. Deliege & J. A. Sloboda (Eds.), *Perception and cognition of music,* (pp. 237-287). Hove, United Kingdom: Taylor and Francis.

Hansen, C. H., & Hansen, R. D. (1988). Finding the face in the crowd: An anger superiority effect. *Journal of Personality and Social Psychology, 54,* 917–924.

Heidemann, K. (2016). A system for describing vocal timbre in popular song. *Journal for the Society for Music Theory, 22*(1). Retrieved from http://www.mtosmt.org/issues/mto.16.22.1/mto.16.22.1.heidemann.html

Hill, B. (2005). *Breaking Down The Breakdown: The Use Of Timbres In Contemporary Dance Music Sub-Genres.* Paper presented at the

Australasian Computer Music Conference 2005, Generate and Test, Queensland University of Technology, Brisbane, Australia.

Houtsma, A. J. M. (1997). Pitch and timbre: Definition, meaning and use. *Journal of New Music Research, 26*(2). doi:10.1080/09298219708570720

Huang, H., & Huang, R. (2013). She Sang as She Spoke: Billie Holiday and Aspects of Speech Intonation and Diction. Jazz Perspectives, 7(3), 287-302. doi:10.1080/17494060.2014.903055

Hughes, R. (2018, February 6). *Janis Joplin: hedonism, heroin, and a life of no half measures.* Retrieved from http://teamrock.com/feature/2018-02-06/janis-joplin-hard-living-hedonism-and-a-life-of-no-half-measures

Huron, D. (2015). The Other Semiotic Legacy of Charles Sanders Peirce: Ethology and Music-Related Emotion. In C. Maeder & M. Reybrouck (Eds.), *Music, Analysis, Experience: New Perspectives in Musical Semiotics* (pp. 185 - 208). Belgium: Leuven University Press.

Huron, D. (2007). *Sweet anticipation: Music and the psychology of expectation.* Cambridge, USA: The MIT Press.

Ihde, D. (2007). *Listening and voice: Phenomenologies of sound* (2nd ed.). Albany, USA: State University of New York Press.

Jairazbhoy, N. A. (1977). The "objective" and subjective view in music transcription. *Ethnomusicology, 21*(2), 263–273.

Johnson-laird, P. N., & Oatley, K. (1992). Basic emotions, rationality, and folk theory. *Cognition and Emotion, 6*(3-4), 201-223. doi:10.1080/02699939208411069

Juslin, P. N. (2001). Communicating emotion in music performance: A review and theoritical framework. In P. N. Juslin & J. A. Sloboda (Eds.), *Music and emotion: Theory and research* (pp. 309–337). Oxford, UK: Oxford Unversity Press.

Juslin, P. N. (2013a). From everyday emotions to aesthetic emotions: Towards a unified theory of musical emotions. *Physics of Life Reviews, 10*, 235–266.

Juslin, P. N. (2013b). What does music express? Basic emotions and beyond. *Frontiers in Psychology. 4*(596). doi:10.3389/fpsyg.2013.00596

Juslin, P. N., & Laukka, P. (2004). Expression, perception, and induction of musical emotions: A review and a questionnaire study of everyday listening. *Journal of New Music Research, 33*(3), 217–238.

Juslin, P. N., & Sloboda, J. A. (2001). Communicating Emotion in Music Performance: A Review and Theoritical Framework. In P. N. Juslin & J. A.

Sloboda (Eds.), *Music and Emotion: Theory and Research* (pp. 309-337). Oxford, United Kingdom: Oxford Unversity Press.

Juslin, P. N., & Timmers, R. (2011). Expression and communication of emotion in music performance. In P. N. Juslin & J. Sloboda (Eds.), *Handbook of music and emotion: Theory, research, applications* (pp. 453–492). Oxford, UK: Oxford University Press.

Katz, M. (2010). *Capturing sound: How technology has changed music*. Berkeley, Los Angeles, USA; London, UK: University of California Press.

Kayser, D. (2016). *Do we feel what we perceive and perceive what we feel? a review of methods in the study of emotions in music* (Unpublished Master's Thesis), University of Oslo, Oslo, Norway.

Keysers, K., & Fadiga, L. (2008). The mirror neuron system: New frontiers. *Social Neuroscience, 3*(3-4), 193-198.

Keysers, C., Kohler, E., Umiltà, M. A., Nanetti, L., Fogassi, L., & Gallese, V. (2003). Audiovisual mirror neurons and action recognition. *Experimental Brain Research, 153*(4). doi:10.1007/s00221-003-1603-5

Koelsch, S., Kasper, E., Sammler, D., Schulze, K., Gunter, T., & Friedreici, A. (2004). Music, language and meaning: brain signatures of semantic processing. *Nature Neuroscience, 7*(3). Retrieved from http://www.cogsci.ucsd.edu/~coulson/CNL/koelsch-needle.pd1f

Kohler, E., Keysers, C., Umilta, M. A., Fogassi, L., Gallese, V., & Rizzolatti, G. (2002). Hearing sounds, understanding actions: Action representation in mirror neurons. *Science, 2*(297), 846-848. doi:10.1126/science.1070311

Krestar, M. L., & McLennan, C. T. (2013). Examining the effects of variation in emotional tone of voice on spoken word recognition. *Q J Exp Psychol (Hove), 66*(9), 1793–1802. doi:10.1080/17470218.2013.766897

Kroeger, K. (1988). Word Painting in the Music of William Billings. *American Music, 6*(1), 41-64.

Lacasse, S. (2000). *"Listen to my voice": The evocative power of vocal staging in recorded rock music and other forms of vocal expression* (Unpublished Doctoral thesis). University of Liverpool, Liverpool, UK.

Lacasse, S. (2010a). The phonographic voice: paralinguistic features and phonographic staging in popular music singing. In A. Bayley (Ed.), *Recorded music: Performance, culture, and technology* (pp. 225–251). New York, USA: Cambridge University Press.

Lacasse, S. (2010b). Slave to the supradiegetic rhythm: A microrhythmic analysis of creaky voice in Sia's 'Breathe Me'. In A. Danielsen (Ed.), Musical

Rhythm in the Age of Digital Reproduction (pp. 141 - 158). London, UK: Routledge.

Lacasse, S. (2010c). The Introspectionist: The Phonographic Staging of Voice in Peter Gabriel's 'Blood of Eden' and 'Digging in the Dirt'. In S. Hill (Ed.), Peter Gabriel, From Genesis to Growing Up (pp. 211 - 224). London, UK: Routledge.

Lacasse, S., & Lefrancois, C. (2008, May). *Intergrating Speech, Music and Sound: Paralinguistic Qualifiers in Popular Music Singing*. Paper presented at the Expressivity in MUsic and Speech, Campinas, Brazil. Retrieved from http://recherche.ircam.fr/equipes/analyse-synthese/EMUS/SP/final_abstracts/EMUS_poster_lefrancois.pdf

Lakoff, G., & Johnsen, M. (2003). *Metaphors we live by*. London, UK: The University of Chicago press.

Lambert, L. C. (1934). *Music Ho! A study of music in decline*. New York, USA: Charles Scribner's Sons.

Lavan, N., Burton, A. M., Scott, S. K., & McGettigan, C. (2018). Flexible voices: Identity perception from variable vocal signals. *Psychonomic Bulletin & Review*. doi:https://doi.org/10.3758/s13423-018-1497-7

Laver, J. (1980). *The phonic description of voice quality*. Cambridge, UK: Cambridge University Press.

Lewicka, M., Czapinski, J., & Peeters, G. (1992). Positive–negative asymmetry or "When the heart needs a reason". *European Journal of Social Psychology, 22*, 425–434.

Lewis, K. M. (2000). When leaders display emotion: How followers respond to negative emotional expression of male and female leaders. *Journal of Organizational Behavior, 21*(2), 221–234.

Lull, J. (1987). Popular music and communication: An introduction. In J. Lull (Ed.), *Popular music and communication* (pp. 10–34). Newbury Park, USA: Sage Publications.

Malkowski, E. F. (2010, 19 September). A new theory for the great pyramid: How science is changing our view of the past. *New Dawn*. Retrieved from http://www.newdawnmagazine.com/articles/a-new-theory-for-the-great-pyramid-how-science-is-changing-our-view-of-the-past

Mann, W. (1963, 27 December). What songs the Beatles sang... *The Times*.

McDonald, J. A. (1967). The "Fish" Cheer/I-feel-like-I'm-fixin'-to-die rag [Recorded by Country Joe and the Fish]. On *I-feel-like-I'm-fixin'-to-die* [vinyl record]. New York, USA: Vanguard.

McDonald Kilmek, M., Obert, K., & Steinhauer, K. (2005). *Estill voice training, Level Two: Figure combinations for six voice qualities*. Pittsburgh, USA: Estill Voice Training Systems International, LLC.

Mellers, W. (1973). *Twilight of the gods*. New York, USA: Schirmer Books.

Merleau-Ponty, M. (1962). *Phenomenology of perception* (C. Smith, Trans.). New York, USA: Routledge and Kegan Paul Ltd. (1945)

Middleton, R. (1993). Popular music analysis and musicology: Bridging the gap. *Popular Music, 12*(2), 177–190.

Middleton, M. (2000). Rock Singing. In J. Porter (Ed.), The Cambridge Companion to Singing (pp. 28 - 41). Cambridge, UK: Cambridge University Press.

Middleton, R. (2000). *Reading pop: Approaches to textual analysis in popular music*. Oxford, USA: Oxford University Press.

Miles, L. K., Stuart, S. B., & Macrae, C. N. (2011, March). *Moving through time*. Paper presented at the 14th Sydney Symposium of Social Psychology: Social Thinking and Interpersonal Behaviour, Sydney, Australia. Retrived from http://www.sydneysymposium.unsw.edu.au/2011/chapters/MacraeSSSP2011.pdf

Missmalevolent. (2007, November 26). *Marilyn Monroe - Happy Birthday Mr. President* [Video file]. Retrieved from https://www.youtube.com/watch?v=EqolSvoWNck

Molnar-Szakacs, I., & Overy, K. (2006). Music and mirror neurons: From motion to "e"motion. *Social Cognitive & Affective Neuroscience, 1*(3), 235–241. Retrieved from http://scan.oxfordjournals.org/content/1/3/235.short

Moore, A. F. (Ed.) (2003). *Analysing popular music*. Cambridge, UK: Cambridge University Press.

Moore, A. F. (2010). The track. In A. Bayley (Ed.), *Recorded music: Performance, culture, and technology* (pp. 252–267). New York, USA: Cambridge University Press.

Moore, A. F. (2012). *Song means: Analysing and interpreting recorded popular song*. Surry, England; Burlington, USA: Ashgate Publishing Limited.

Moylan, W. (2007). *Understanding and Creating the Mix: The Art of Recording*. Burlington: Focal Press.

Nattiez, J. (1990). *Music and discourse: Towards a semiology of music*. New Jersey, USA: Princeton University Press.

Neal, J. (2000). Songwriter's Signature, Artist's Imprint: The Metric Structure of a Country Song. In C. K. Wolfe & J. E. Akenson (Eds.), *Country Music Annual 2000* (pp. 112-140.). Lexington, USA: The University Press of Kentucky.

Neal, J. (2007). Narrative paradigms, musical signifiers, and form as function in country music. *Music Theory Spectrum, 29*(1), 41–72.

Nygaard, L. C., & Lunders, E. R. (2002). Resolution of lexical ambiguity by emotional tone of voice. *Memory and Cognition, 30*(4), 583–593.

Nygaard, L. C., & Quees, J. S. (2008). Communicating emotion: Linking affective prosody and word meaning. *Journal of Experimental Psychology: Human Perception and Performance, 34*(4), 1017–1030.

Ortony, A., & Turner, T. J. (1990). What's basic about basic emotions? *Psychological Review, 97*(3), 315-331. doi:http://dx.doi.org/10.1037/0033-295X.97.3.315

Paul, B., & Huron, D. (2010). An Association between Breaking Voice and Grief-related Lyrics in Country Music. *Empirical Musicology Review, 5*(2), 27 - 35.

Piazza, E. A., Cătălin Iordan, M., & Lew-Williams, C. (2017). Mothers consistently alter their unique vocal fingerprints when communicating with infants. *Current Biology, 27*(20), 3162-3167. doi:https://doi.org/10.1016/j.cub.2017.08.074

Plutchik, R. (1991). *The Emotions*. Maryland: University Press of America, Inc.

Porter, C. (1956). I've got you under my skin [Recorded by Frank Sinatra]. On *Songs for swingin' lovers* [CD]. California, USA: Capitol. (1987)

Poulin-Charronnat, B., Bigand, E., Madurell, F., & Peereman, R. (2005). Musical structure modulates semantic priming in vocal music. *Cognition, 94*(3), B67-78. doi:10.1016/j.cognition.2004.05.003

Poyatos, F. (1992). The Audible-Visual Approach to Speech as Basic to Nonverbal Communication Research. In F. Poyatos (Ed.), *Advances in Nonverbal Communication: Sociocultural, Clinical, Esthetic and Literary Perspectives* (pp. 41 - 58). Amsterdam: John Benjamins Publishing Company.

Poyatos, F. (1993). Paralanguage: A linguistic and interdisciplinary approach to interactive speech and sound. Amsterdam: John Benjamins.

Poyatos, F. (2002). *Nonverbal communication across disciplines: Volume II: Paralanguage, kinesics, silence, personal and environmental interaction* (Vol. II). Amsterdam, The Netherlands; Philadelphia, USA: John Benjamins B.V.

Poyatos, F. (1975). Cross-Cultural Study of Paralinguistic "Alternants" in Face-to-Face Interactions. In A. Kendon, R. M. Harris, & M. R. Key (Eds.), *Organization of Behavior in Face-to-Face Interaction* (pp. 285 - 314). Netherlands: Mouton Publishers.

Proverbio, A. M., Riva, F., & Zani, A. (2009). Observation of static pictures of dynamic actions enhances the activity of movement-related brain areas. *PLOS ONE, 4*(5). doi:10.1371/journal.pone.0005389

Qualtrics. (2005). *Qualtrics* [Computer software]. Retrieved from http://www.qualtrics.com

Ragna. (2012, July 25*). West Side Story - 14. Somewhere* [Video file]. Retrieved from https://www.youtube.com/watch?v=g3mZY8mf2hU

Pasler, J. (2001, January 01). Impressionism. Grove Music Online. Ed. Retrieved 24 Aug. 2018, from http:////www.oxfordmusiconline.com/grovemusic/view/10.1093/gmo/9781561592630.001.0001/omo-9781561592630-e-0000050026.

Reed, S. A. (2005). *The musical semiotics of timbre in the human voice and static takes love's body*. (Unpublished Doctoral thesis). University of Pittsburgh, Pittsburgh, USA. Retrieved from http://d-scholarship.pitt.edu/7313/1/s.a.reed_2005etd.pdf

Reevy, G., Ozer, Y. M., & Ito, Y. (2010). *Encyclopedia of emotion*. Santa Barbara, USA: Greenwood.

Reznikoff, I. (1995). On the sound dimension of prehistoric painted caves and rocks. In E. Tarasti (Ed.), *Musical signification: Essays in the semiotics theory and analysis of music* (pp. 541–557). Berlin, Germany: Mouton De Gruyter.

Rinkenauer, G., Osman, A., Ulrich, R., Müller-Gethmann, H., & Mattes, S. (2004). On the locus of speed-accuracy trade-off in reaction time: Inferences from the lateralized readiness potential. *Journal of Experimental Psychology: General, 133*(2), 261–282. doi:http://dx.doi.org/10.1037/0096-3445.133.2.261

Rodet, X., Potard, Y., & Barrière, J.-B. (1984). The CHANT project: From the synthesis of the singing voice to synthesis in general. *Computer Music Journal, 8*(3), 15–31.

Rossing, T. D. (1990). *The Science of Sound*. Reading, Massachusetts: Addison-Wesley Publishing Company.

Rozin, P., & Royzman, E. (2001). Negativity bias, negativity dominance, and contagion. *Personality and Social Psychology Review, 5*(4), 296–320.

Russell, J. A., & Fernández-Dols, J. M. (1997). What does a facial expression mean? In J. A. Russell & J. M. Fernández-Dols (Eds.), *The psychology of facial expression* (pp. 3–30). Cambridge, UK: Cambridge University Press.

Sabine, W. C. (1922). *Collected papers on acoustics*. Cambridge, USA: Harvard University Press.

Sapp, C. (2016). Suggestions for Future Corpus-Based Text Painting Analyses: A Response to Strykowski. *Empirical Musicology Review, 11*(2).

Schaeffer, P. (1966). *Traite des objets musicaux.* Paris, France: Seuil.

Schirmer, A., & Kotz, S. A. (2003). ERP evidence for a sex-specific stroop effect in emotional speech. *Journal of Cognitive Neuroscience, 15*(8), 1135–1148.

Seashore, C. E. (1938). The Psychology of Music New York: McGraw Hill Book Company Inc.

Seeger, C. (1977). *Studies in musicology 1935- 1975.* Berkeley, USA; London, UK: Univeristy of California Press.

Shadoian, J. (1971, 18 February). *Janis Joplin: Pearl.* [*Review of the vinyl record Pearl, by Janis Joplin,* Columbia, 1971]. *Rolling Stone*. Retrieved from https://www.rollingstone.com/music/albumreviews/pearl-19710218

Slepian, M. L., Weisbuch, M., Rule, N. O., & Ambady, N. (2011). Tough and tender: Embodied categorization of gender. *Psychological Science, 22*(1), 26–28. doi:10.1177/0956797610390388

Smalley, D. (1986). Spectro-morphology and structuring processes. In S. Emmerson (Ed.), *The language of electroacoustic music*. Hong Kong, Hong Kong: The McMillian Press Ltd.

Smalley, D. (1994). Defining timbre, refining timbre. *Contemporary Music Review, 10*(2), pp. 35–48.

Smalley, D. (1997). Spectromorphology: explaining sound-shapes. *Organised Sound, 2*(2), 107–126.

Smith, A. (2011). *The performance of 16th-century music: Learning from the theorists*. New York, USA: Oxford University Press.

Smith, N. A., & Trainor, L. J. (2008). Infant-directed speech is modulated by infant feedback. *Infancy, 13,* 410–420.

Soto, D., Funes, M. J., Guzman-Garcia, A., Warbrick, T., Rotshtein, P., & Humphreys, G. W. (2009). Pleasant music overcomes the loss of awareness in patients with visual neglect. *Proceedings of the National Academy of Sciences, USA, 106*(14), 6011–6016. doi:10.1073/pnas.0811681106

Stevens, K. N. (2000). *Acoustic Phonetics*. Cambridge, Massachusetts: MIT Press.

Strykowski, D. R. (2016). Text Painting, or Coincidence? Treatment of Height-Related Imagery in the Madrigals of Luca Marenzio. *Empirical Musicology Review, 11*(2).

Sundberg, J., & Thalen, M. (2010). Journal of Voice. *What is "Twang"?, 24*(6), 654-660.

Sundberg, J. (1997). The Acoustics of the Singing Voice. *Scientific American, 236*(3), 82-91.

Tagg, P. (1999). *Introductory notes to the semiotics of music*. Retrieved from https://www.tagg.org/xpdfs/semiotug.pdf

Tagg, P. (2000). Analysing popular music: theory, method and practice. In R. Middleton (Ed.), *Reading pop: Approaches to textual analysis in popular music*. Oxford, UK: Oxford Unversity Press.

Tagg, P. (2011, May). *Music, moving image and the "missing majority"*. Paper presented at the Music & The Moving Image, Steinhardt School, New York University, New York, USA. Retrieved from https://tagg.org/articles/xpdfs/NYC110521.pdf

Tagg, P. (2012). *Music's Meanings* (version 2.5.2 ed.). (n.p.): Author.

Tagg, P., & Clarida, B. (2003). *Ten little title tunes: Towards a musicology of the mass media*. New York, USA; Montreal, Canada: Mass Media Music Scholars Press.

Tartter, V. C. (1980). Happy talk: Perceptual and acoustic effects of smiling on speech. *Perception & Psychophysics, 27*(1), 24–27.

Tartter, V. C., & Braun, D. (1994). Hearing smiles and frowns in normal and whisper registers. *The Journal of the Acoustical Society of America, 96*, 2101–2107. doi:http://dx.doi.org/10.1121/1.410151

Telestream. (2016). *ScreenFlow* [Computer software]. Retrieved from http://www.u.arizona.edu/~jforster/dmdx/

Temperley, D. (1999). Syncopation in rock: A perceptual perspective. *Popular Music, 18*(1), 19–40.

Temperley, D. (2007). The melodic-harmonic "divorce" in rock. *Popular Music, 26*(2), 323–342.

Tillmann, B., & Bigand, E. (2002). A comparative review of priming effects in language and music. In P. Mc Kevitt, S. O. Nuallain, & C. Mulvihill (Eds.), *Language, vision and music: Selected papers from the 8th International Workshop on the Cognitive Science of Natural Language Processing, Galway, Ireland, 1999* (pp. 231-241). Amsterdam, The Netherlands: John Benjamins Publishing Co.

Titze, I. R. (1989). Physiologic and acoustic differences between male and female voices. *The Journal of the Acoustical Society of America, 85*(4), 1699-1707.

van Leeuwen, T. (1999). *Speech, music, sound*. London, UK: Palgrave Macmillian.

van Leeuwen, T. (2012). The critical analysis of musical discourse. *Critical Discourse Studies, 9*(4), 319–328.

Walser, R. (1995). Rhythm, rhyme, and rhetoric in the music of Public Enemy. *Ethnomusicology, 39*(2), 193–217.

Warrier, C. M., & Zatorre, R. J. (2002). Influence of tonal context and timbral variation on perception of pitch. *Perception & Psychophysics, 64*(2), 198–207.

Waterman, D. (2012). The Story Behind The Song: Adele, "Someone Like You".   Retrieved from http://americansongwriter.com/2012/01/the-story-behind-the-song-adele-someone-like-you/

Wescott, R. W. (1992). Auditory Communication: Non-Verbal, Pre-Verbal, and Co-Verbal. In F. Poyatos (Ed.), *Advances in Nonverbal Communication: Sociocultural, Clinical, Esthetic and Literary Perspectives* (pp. 25 - 40). Amsterdam: John Benjamins Publishing Company.

Wess, P., & Taruskin, R. (2008). *Music in the western world: A history in documents* (2nd ed.). Belmont, USA: Thomson Schirmer.

Wickelgren, W. (1977). Speed-accuracy tradeoff and information processing dynamics. *Acta Psychologica, 41*(1), 67–85. doi:http://dx.doi.org/10.1016/0001-6918(77)90012-9

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review, 9*(4), 625–636. doi:10.3758/BF03196322

Wilson, M. (2011). *Examining the effects of variation in emotional tone of voice on spoken word recognition.* (Unpublished Master's Thesis), Cleveland State University, Ohio, USA. Retrived from https://engagedscholarship.csuohio.edu/etdarchive/569/

Winkler, P. (1997). Writing ghost notes: The poetics and politics of transcription. In D. Schawarz, A. Kassabian, & L. Siegel (Eds.), *Keeping score: Music, disciplinarity, culture* (pp. 169–203). Charlottesville, USA; London, UK: University Press of Virginia.

Wood Massi, R. (1992). *Music and discourse: Toward a semiology of music by Jean-Jacques Nattiez, Carolyn Abbate*. [Review of the book Music and discourse, by J. Nattiez & C. Abbate, Trans.]. *Notes, 48*(4), 1286–1288. Retrieved from http://www.jstor.org/stable/942137

Wright, R. (2006, May). *Intra-speaker variation and units in human speech perception and ASR*. Paper presented at the International Speech Communication Association Conference ITRW on Speech Recognition and Intrinsic Variation, Toulouse, France. Retrived from http://www.isca-speech.org/archive_open/archive_papers/sriv2006/sriv_039.pdf

Zacher, V., & Niemitz, C. (2003). Why can a smile be heard? A new hypothesis on the evolution of sexual behaviour and voice. *Anthropologie, 41*(93 -98).

Zagorski-Thomas, S. (2014). *The Musicology of Record Production*. Cambridge: Cambridge University Press.

Zak III, A. J. (2001). *The Poetics of Rock: Cutting Tracks, Making Records*. Berkeley; Los Angeles; London: Univeristy of California Press.

Zarate, J. M., Xing Tian, K., Woods, J. P., & Poeppel, D. (2015). Multiple levels of linguistic and paralinguistic features contribute to voice recognition. *Nature, 5*. doi:http://dx.doi.org/10.1038/srep11475

Zarate, J. M., Ritson, C. R., & Poeppel, D. (2013). The effect of instrumental timbre on interval discrimination. *PLOS ONE, 8*(9). doi:10.1371/journal.pone.0075410

Zbikowski, L. M. (2009). Music, language, and multimodal metaphor. In C. Forceville & E. Urios-Aparisi (Eds.), Multimodal Metaphor. Berlin, Boston: De Gruyter Mouton.