

EMOTIONAL TONES AND EMOTIONAL TEXTS:

A NEW APPROACH TO ANALYSING THE VOICE IN POPULAR VOCAL SONG.

This paper in its current form has been accepted for publication in Music Theory Online, <https://www.mtosmt.org/index.php>.

It is scheduled for publication in Volume 28, no. 2, June, 2022.

This document will be updated with the preferred citation for the published work in June 2022. In the meantime, please cite this work as follows:

Spreadborough, K. (2021). *Emotional tones and emotional texts: A new approach to analysing the voice in popular vocal song*. University of Melbourne Figshare. <https://doi.org/10.26188/16900318>

Abstract: Vocal tone quality is a highly emotive musical resource in popular vocal songs. However, it is also one of the most difficult aspects to analyse due to the complexity and variety of the voice. This paper presents a novel analytical approach to the sung voice by considering how emotion is conveyed through tone quality and text. The aim of the approach is to provide a system for annotating vocal tone quality and for analysing its emotive content. The approach is informed by findings from psychology, music studies, and the social semiotics of sound – taking into consideration how our everyday experience of voice in communication contributes to our emotional perception of singing. Different modes of annotation, from static annotation to real-time annotation, are demonstrated. This paper first presents the theoretical underpinnings of the approach, followed by an outline of the approach itself, and finally demonstrates the approach through an analysis of the vocal line in Kris Kristofferson’s 1970 song “Casey’s Last Ride”.

Key words: Voice quality, tone quality, emotion, song, popular music, music psychology, social semiotics, analysis

1 Introduction

[1]Humans have a rich pallet of vocal cues to express meaning, from literal linguistic meanings to more abstract emotional expression (see, for example, Poyatos 1993). Research has shown that similar cues can also play a role in musical communication (Juslin and Laukka 2003). The sung voice is one area in which the connection between spoken and musical communication is most apparent. This is especially the case in popular vocal songs, where an artist's spoken voice often plays a central role in their vocal aesthetic (Lacasse 2010, 142). Because of this speaking-singing connection, the lived experience of speaking may heighten one's sensitivity to vocal expression in popular vocal songs. As Simon Frith observes

The voice is a direct expression of the body, ... it 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, 192)

[2]Given its near ubiquity in much of popular music, and its potential for emotional expression, the analysis of voice in popular vocal songs is a fruitful avenue of research. However, there are few techniques which allow such an analysis (Author 2018, chap. 4). One reason for this is the complexity of vocal sounds that occur within and between popular vocal songs. On a physiological level, this complexity is due to no two human bodies being exactly

alike. Basic biological differences can impact the size and shape of these structures (Titze 1989), thus impacting one's overall vocal quality. 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 quality (Lavan et al. 2018, Introduction). Furthermore, a single human vocal tract can be manipulated to produce a multitude of different vocal qualities (Lavan et al. 2018). Consider, for example, the differences in commonly heard vocalisations such as laughter, whispering, shouting, and speaking. Thus, the sung voice presents a unique challenge for music analysis: How, within this rich spectrum of acoustic cues that differ both within and between performers and performances, is one to achieve a coherent and systematic analysis of the sung voice?

[3] This paper takes a social semiotic approach to the analysis of tone quality in song. Social semiotics considers "what you can 'say' with *sound*, and how you can interpret the things that other people 'say with sound'" (van Leeuwen 1999, p. 4). The social semiotic approach places our perception and lived experience of sound at the centre of our understanding of sound. Others have sought to describe and analyse the sound characteristics of popular and recorded music; however none have yet drawn on social semiotics to develop a systematic framework for the analysis of voice quality specifically. Work in electroacoustic and recorded music has faced similar challenges in describing and analysing acousmatic sounds that are not traditionally notated (e.g. Smalley 1986, 1997; Moylan 2015). Some of these works have drawn on a similar set of terminology as will be used in this paper (e.g. Moylan's Physical Dimensions, Perceived Parameters, and Artistic/Aesthetic Elements, see Moylan 2015, Chapter 2). However, such approaches do not analyse music through a social semiotic lens. Other approaches have sought to define and examine specific characteristics of sound (e.g. Lomax 1968), or through mechanisms to describe in detail the various sounds employed in vocal production (e.g. Wishart 1996, Chapter 12). However such work does not provide a

mechanism through which the social semiotics of the voice can be considered, nor do they provide a structured framework for annotating and analysing the sung voice.

[4]In this paper I address this gap by proposing a new approach to analysing the vocal line in popular vocal songs that is grounded in the social semiotics of sound, and draws on music psychology and music studies more broadly. The focus of this approach is to provide a system through which to annotate and analyse the sound of the sung voice (hence forth referred to as tone quality) in terms of its emotive content. This approach also allows the emotive content of tone quality to be compared with that of text (i.e. lyrical content), and for the implications of this relationship on overall emotional perception of the vocal line to be assessed.

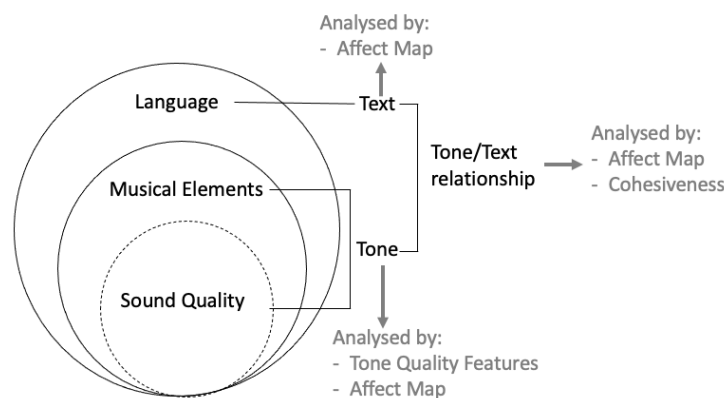
1.1 The Analytical Approach

[5]Example 1 shows a high-level conceptualisation of the analytical approach. The text in grey indicates the tools available for use at each level of the annotation and/or analysis. This conceptualisation is based on the premise that singers perform (*sound quality*) music (*musical elements*) with lyrics (*text*). Since all musical sounds must be produced by an instrument, and each instrument has its own unique sound quality, *sound quality* is at the heart of the conceptualisation. In the case of popular vocal songs, this is even more apparent as idiosyncratic vocal sound qualities are often central to artist's performance style (Heidemann 2016, 2). At the next level are *musical elements*, which are those features that are more commonly associated with traditional musical analysis (e.g. pitch, dynamic). Sound quality and musical elements together constitute tone quality.

[6]The broken line between sound quality and musical elements demonstrates that while these are definable attributes of a tone quality, these are not discrete, wholly separate layers. Rather, they are dialectically related (Fairclough 2001, 234). That is, sound qualities may play a role in musical elements, and vice versa. This is discussed in more detail in Section

3.2.2 below, however it is worth mentioning here since it is due to this dialectic relationship between sound quality and musical elements that tone quality more broadly is taken as the focus of this analytical technique. Tone quality is annotated and analysed using the **tone quality features** discussed in Section 3.2. The emotionality of a tone quality is assessed against the **Affect Map** discussed in more detail in Section 2.1.

[7]*Text* (i.e. lyrical content) is at the outermost layer in the conceptualisation. Text includes both words and non-words that accompany the vocal performance. The approach to analysing text is discussed in Section 4. Emotionality of text is assessed against the **Affect Map** mentioned above. The tone/text relationship is the part of the conceptualisation where the emotionality of tones and texts are compared and contrasted. This is achieved through the use of the tool **Cohesiveness**, which is discussed in more detail in Section 2.2.



Example 1. A high-level conceptualisation of the analytical approach. (The figure is the authors original work).

1.2 Lived Experience, Emotion Perception, and Popular Vocal Song Analysis

[8]The emotional models which underpin this analytical approach will be discussed in more detail in Section 2.1.1 below. Before undertaking this discussion, however, it is useful to explore in more detail how our lived experience of speaking informs our emotional

perception of the sung voice, and to consider other analytical approaches which have drawn on this connection in analysing popular vocal songs.

[9]In spoken word contexts, paralinguistic features, the “non-phonemic alterations of the pitch, stress, or tempo of ordinary speech, as in growling, shouting, or drawling” (Wescott 1992, 30), are important for optimal verbal communication (Wilson 2011). Fernando Poyatos observed that “words lack the capacity to carry the whole weight of a conversation, as our verbal lexicons are extremely poor in comparison with their 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). Paralinguistic features of speech have the power to emphasize a message, “deemphasize it or contradict it altogether” (Poyatos 1992, 51).

[10]Similar features have been found to play a key role in musical expression. For example, it has been found that the “breaking voice” plays an essential role in conveying grief in Country songs (Paul and Huron 2010). Similarities of tone quality in music and paralinguistic cues in speech have also been found in the expression of sarcasm, which appears to be reliably marked by nasality in both spoken and musical contexts (Palzak 2010, as cited in Huron 2015, 190). Parallels have also been found between the paralinguistic and tone quality expression of sadness where sadness tends to be conveyed through darker timbres in both spoken and musical contexts (Huron 2015, 193). Such research suggests that drawing on one’s experience of spoken voice is a fruitful avenue for analysing emotive content of the sung voice.

[11]Indeed, previous musical analyses have drawn on this speaking-singing connection. For example, Richard Middleton outlines how rock singing can be viewed as a spectrum where, at one extreme words govern the song, and the voice tends towards speech as it delivers the narrative (Middleton 2000, 29). At the other extreme, words “are absorbed into the musical flow, working as sound or gesture”, and the voice becomes an instrument (Middleton 2000,

30). Singing, viewed as a stylised form of speech, may sit anywhere along this spectrum. Serge Lacasse too explores this connection. One example is his 2010 analysis of Sia's "Breath Me". Here, Lacasse approached the sung voice as "a stylised means of conveying emotion using, among other things, paralinguistic features borrowed from everyday speech" (Lacasse 2010, 142). Lacasse outlines several levels on which links between the spoken and sung voice can be examined and explores how such emotional utterance in voice may play a key role in conveying emotion in song (Lacasse 2010, 143).

[12] Despite these and other similar approaches to vocal song analysis (e.g., Heidemann 2016; van Leeuwen 1999), no technique yet offers a systematic approach to annotating and analysing tone quality in the sung voice, especially in terms of emotion perception. This is the goal of the present paper. This paper will first define the parameters of analysis through a discussion of the emotional models which underpin the analysis and outline the tools for discussing emotion – the Affect Map and Cohesiveness. Next, the system for annotating and analysing tone quality, the tone quality features, will be presented. Following this, an overview of the approach to textual analysis will be given. Finally, the analytical approach will be applied to a case study, "Casey's Las Ride", to demonstrate the potential for application.

2 Defining the Parameters of Analysis

[13] This section will outline the two tools used to discuss emotion within and between tone quality and text: the Affect Map and Cohesiveness (see Example 1). The Affect Map will be discussed first because it includes a discussion of the emotional models which underpin this paper. The tool Cohesiveness will then be described. These tools are inspired by diagrammatic vocabulary sets, an approach developed as part of Denis Smalley's analytical techniques for electroacoustic music (e.g., Smalley 1986, 1997).

2.1 The Affect Map

[14]Before emotion per se is examined, it is important to consider the *locus* of emotion. This paper draws on the locus of emotion as described by Evans and Schubert (2008). This model is used since it provides an account of the relationship between loci of emotion that is grounded within music psychology – one of the disciplines that informs the approach proposed in this paper. According to Evans and Schubert, the locus of emotion refers to how listeners experience emotion in music – either by recognising emotion expressed by the music, called perceived or external locus of emotion; or through feeling a subjective response to the music, called felt or internal locus of emotion. External and internal loci of emotion interact in a number of ways, but the mechanisms behind this interaction and the attribution of emotion to the external or internal loci require further investigation (Evans and Schubert 2008). In this paper, I will focus on the external locus of emotion, discussing within the analysis emotions which may be *perceived* by the listener.

[15]Two main models tend to be used to understand emotion in music: the discrete model and the dimensional model (Eerola and Vuoskoski 2013, 317). The discrete model tends to focus on basic emotions, a small set of “evolutionary emotions that have important functions when adapting the individual to events that have material consequences for the individual’s well-being” (Eerola and Vuoskoski 2013, 310). The most common dimensional model is that of Russell (1980). It considers how emotion can be understood through two systems, valence (pleasure-displeasure) and arousal (activation-deactivation) (T. Eerola and Vuoskoski 2011). Another dimensional mode (e.g., Tellegen, Watson, and Clark 1999) involves considering emotion in terms of arousal–calmness and tension–relaxation, and uses these systems to infer valence (Eerola and Vuoskoski 2011).

[16]The dimensional models discussed above examine emotion from two dimensions only. However, some have argued that emotion is better understood through three dimensions. In

their 2000 paper, Schimmack and Grob proposed a model with three dimensions: tension arousal (tense-relaxed), energy arousal (awake-tired) and valence (pleasant-unpleasant). The similarities and differences of the discrete model and Schimmack and Grob's (2000) three-dimensional model was explored by Eerola and Vuoskokski (2011). It was found that participants were able to consistently accurately rate emotional musical stimuli on both the discrete and dimensional models, suggesting that individuals can understand emotions well both in terms of basic emotions (happy, sad, etc) and descriptors on a spectrum (very pleasant, moderately pleasant, etc). However, they also found that participants were able to more accurately rate ambiguous emotions when using the dimensional model. The authors suggest that a hybrid model may be useful for music emotions research. Such a model

... uses the components of a dimensional model (valence and arousal) to explain the underlying affect space, which is mainly physiologically driven. When the changes in these core affects are interpreted consciously, however, discrete emotion terminology is used to label the emotional experiences. In this way common discrete emotions can be regarded as attractors or hot spots in the affect space. (Eerola and Vuoskokski 2011, 41)

[17]Although the discussion thus far has been focused on emotion, it is also sometimes necessary to discuss mood within music. Unlike emotions, which tend to be short lived affective states and emotion words tend to “imply an object (e.g. I love somebody, I am afraid of dogs)” (Schimmack and Grob 2000, 328), moods are longer lived and mood words “are not directed at objects (e.g. I am relaxed, I am tired)” (Schimmack and Grob 2000, 328). There is much debate in current music psychology literature about how music evokes moods, what kind of moods are evoked, and indeed where the boundary between emotion and mood lies in musical experience (Hunter and Schellenberg 2010, section 5.2). However, it is generally agreed that regardless of the exact mechanisms and classifications, music does

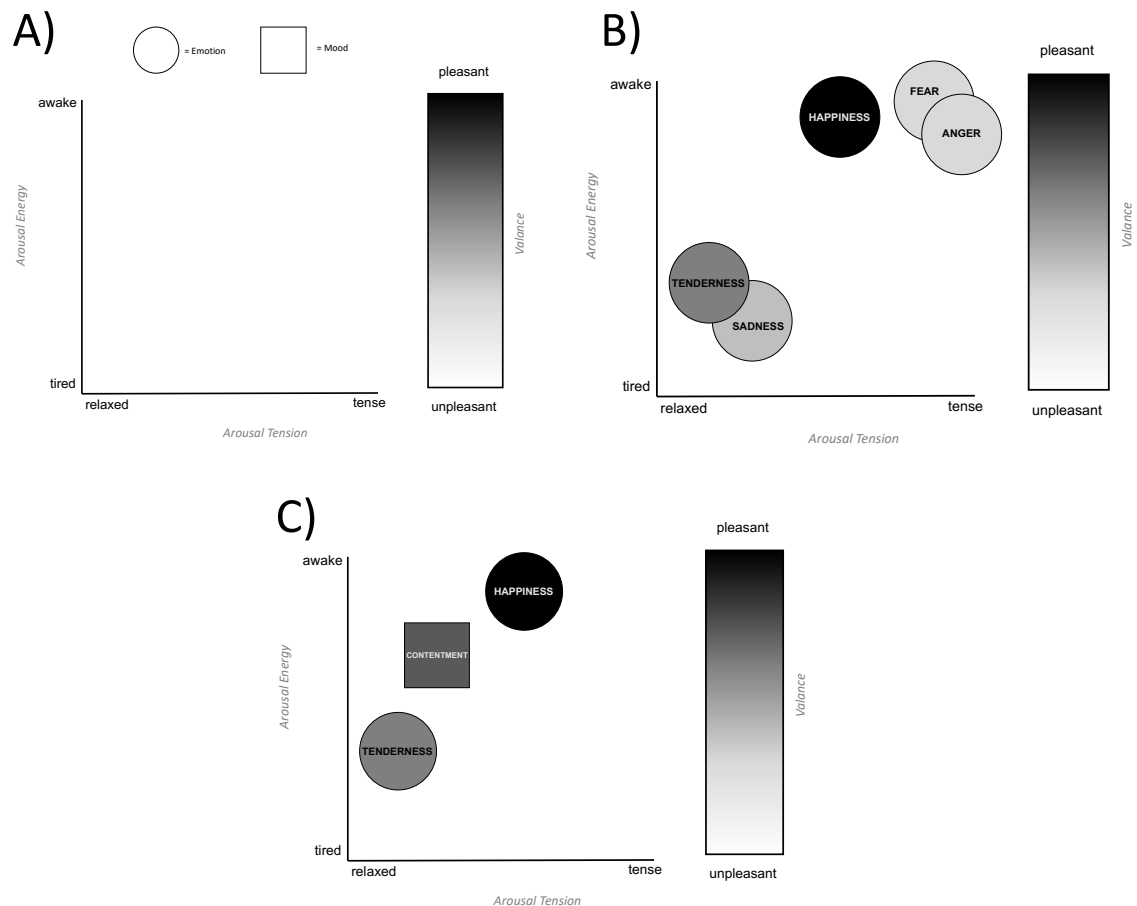
induce mood and mood is a more diffuse experience that is typically not directed at a target (as is the case for emotion) (Hunter and Schellenberg 2010, section 5.2). To account for both diffuse moods (e.g. when commenting on the mood of an entire musical section) and targeted emotions (e.g. when commenting on the emotion conveyed through specific tone quality features in song), I use the “Affect Map” (Example 2).

[18]The term affect is used since it typically refers to overall reaction to music that encompasses both moods and emotions. In music specifically, affect is used as a term for describing individuals’ responses to sound, although there is debate about how mood and emotion contribute to musical affect (Hunter and Schellenberg 2010, section 5.8). Since I wish to refer to both moods and emotions in my analysis, the term Affect has been adopted here to describe the relationship between the two. No predefined taxonomy of emotions or moods is used in the Affect Map for two reasons. First, there is much debate in the literature about what kinds of moods and emotions music can evoke and convey (Hunter and Schellenberg 2010). Therefore, there is no widely agreed upon taxonomy on which to draw from the music psychology literature. Second, the Affect Map is not intended to provide a taxonomy of moods and emotions. It is designed to provide a framework through which one can document affect in tone quality analysis. Not tying the framework to a taxonomy, therefore, provides the flexibility of the framework to be applied in a descriptive way across a number of different contexts.

[19]The Affect Map provides a visual, dynamic representation of emotion and mood in light of the literature discussed above. A blank template is shown in Example 2A. Moods are denoted by squares and emotions are denoted by circles. Valence is denoted by the colour which sits on a spectrum from white to black, and arousal tension and arousal energy lay along the x and y axis respectively. The Affect Map can be used to represent any number of emotions, as shown in Example 2B. The placement of the emotions along the arousal and

valance scales are my own. They are intended to be indicative only – a demonstration of how emotions can be placed on the Affect Map. There are of course different levels of intensity of each emotion and the placement of emotions on the Affect Map can vary to demonstrate the varying levels of intensity. The placement of moods (demonstrated in example 2C) can also vary along the scale. The placement of moods is also my own and derived from my personal experiences.

[20]Example 2C shows how the Affect Map may be used – in this example it is used to represent emotions of happiness and tenderness, thus creating a mood of contentment. Happiness in this example is positive (pleasant valance), of moderate energy (moderate arousal tension) and quite alert (high arousal energy). Tenderness in this example is moderately positive (moderately pleasant valance), relaxed (low arousal tension) and quite sedate (lower arousal energy). Taken together, these two emotions create a mood of contentment – moderate valance, slightly elevated arousal energy, and low arousal tension. Of course, happiness and tenderness may be placed at different points on the three scales (there are different kinds of happiness and tenderness). The examples used here are intended to be illustrative only of how the Affect Map can be used, not prescriptive representations of the level of valance, arousal energy, and arousal tension that make up each mood and emotion. How tone quality features in their extreme form relates to the Affect Map is outlined in the tables in section 3.2 below.



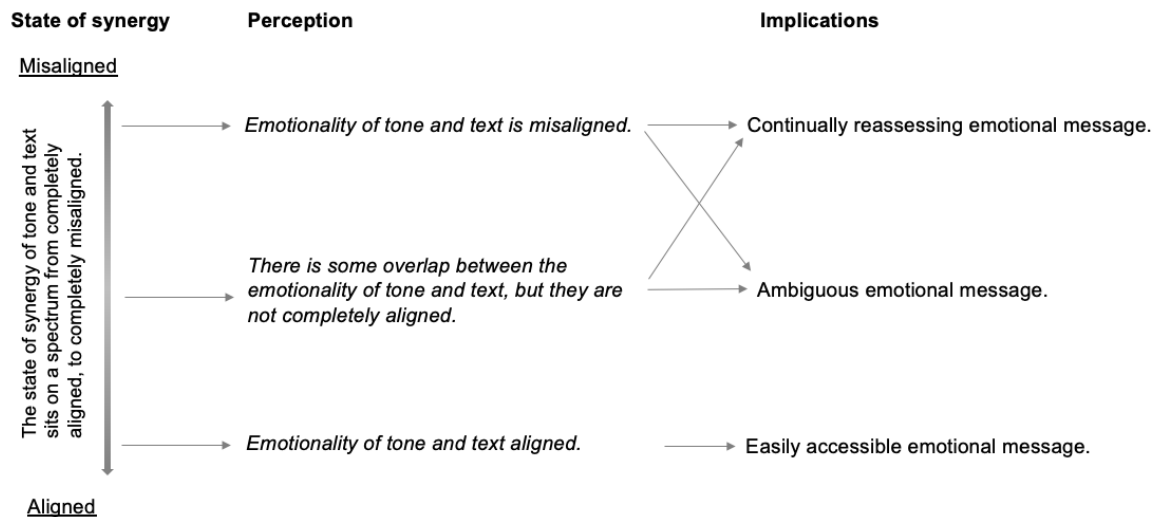
Example 2. The Affect Map, a graphical representation of moods and emotions represented first through the dimensional model of emotions, then assigned discrete labels based on this placement. (The figure is the authors original work).

2.2 Cohesiveness

[21]Cohesiveness, shown in Example 3 below, is a tool developed for this analytical technique. It allows the analyser to explore the synergies and conflicts in emotion conveyed through tone and text, and to consider the implications of this for the vocal line in general. In examining the relationship between tone and text, this work is situated within a long tradition of research on the interaction of words and music. Of particular relevance to this paper is the concept 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). As metaphor, word painting can be used to convey, subvert, or strengthen a lyrical message (see, for example, Kroeger, 1988; 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 also been explored. For example, Serge Lacasse has explored how “the manipulation of voice through recording techniques can contribute to the mediation of... expressive moments” (Lacasse, 2010c, p. 211). The analytical technique proposed in the present paper draws on this long tradition of music-text analysis by not only presenting a novel method for analysing music (tone quality) but by also proposing a systematic mechanism through which one can examine both tone and text together – Cohesiveness.

[22]Cohesiveness is a tool to allow the simultaneous assessment vocal and linguistic expression. 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”, one may identify the tone quality as being positive 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 mood. The emotional message is ambiguous, and the listener must continually reassess the tone and text to determine what message is being conveyed by the performer. The application of Cohesiveness will be demonstrated in section 5 below.



Example 3. Cohesiveness. A tool to describe the state of synergy between the emotionality of tone and text and the flow-on effects for emotion perception. (The figure is the authors original work).

3 Annotating and Analysing Tone

[23]This section presents a classification system, called the tone quality features, for annotating and analysing certain acoustic cues within a tone quality. The features identified for inclusion in the framework are based on a social semiotic perspective of voice quality. This perspective views the emotive power of voice quality as arising from the configuration of different vocal dimensions (e.g., a voice is never just low, but it is also smooth and soft) (van Leeuwen 1999, p. 129). It also emphasises the examination of sound not only in terms of what it “expresses” or “represents”, but also how it “affects us” (van Leeuwen 1999, p. 128). The tone quality features identified here are drawn on from existing social semiotic approaches to voice quality (van Leeuwen, 1999, Ngo et al., 2021). I have extended on these features by including a system for describing onsets, something which is not addressed in existing social semiotic approaches to voice quality, and expanding on the emotional

implications for each tone quality feature. I have also developed a system for annotating the tone quality features. Each feature has been assigned a unique symbol which can be used in place of the linguistic description (for example, a breathy sustain is symbolised as $\rightarrow\rightarrow$)¹.

Benefits of the tone quality features

[24]Being able to succinctly and consistently describe acoustic features of a tone quality has two benefits. First, 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 and discussing such elements as pitch and rhythm. Having a system to describe aspects of a tone quality, then, also lends tone quality analysis a level of clarity which may not otherwise be available.

[25]Second, it may help to identify emotional valence. Having a system for discussing tone quality is the first step in being able to discuss how emotion is conveyed through that tone

¹ Some labels used in the tone quality features 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 used here and 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 tone quality will be determined exclusively by the properties of the phoneme being sung. The potential overlap between the phonetic requirements and the tone quality features will be taken into account in the analyses. Additionally, and in line with the social semiotic approach taken here, analysis is not undertaken at the level of the phoneme but at the level of the word and phrase (see Section 3.1).

quality. For example, some research has suggested that the form of tone quality features may be associated with particular emotions, such as roughness and negative emotions (e.g. Author 2018, Chapter 6), due to our lived experience of sounds (as discussed in Section 1.2 above). Before this link can be made, however, a systematic way of annotating and discussing these features is required. This is the goal of the tone quality features.

The relationship between sound quality and musical elements

[26]As can be seen in the subheadings below, there are a number of features which make up tone quality. Example 1, presented in the introduction, shows sound quality and musical elements, which constitute tone quality, separated by a broken line. This represents that musical elements and sound quality are not discrete, but they are dialectically related (Fairclough 2001, 234). For example, some of the features listed below constitute musical elements, such as dynamic, and dynamic is different to a feature which constitutes sound quality, such as breath. However dynamic and breath, and by extension too musical elements and sound quality, remain in dialogue in that it is easier to produce more breathy sounds at softer dynamics than at louder ones. It is for this reason that the features below are not categorised into discrete, separate lists for sound quality and musical elements – because each feature, while being discernible and definable in its own right, is in constant dialogue with other features and by extension so too are musical elements and sound qualities.

[27]In this paper, tone quality features which could be considered musical elements have only been included where they can also contribute to sound quality (namely, dynamic, range, and vibrato). Thus, in the tone quality features below, not all sound qualities are musical elements, but all musical elements contribute to sound qualities. It is for this reason that other musical elements such as tempo, duration, and range are not included in this analytical approach – because they do not contribute to sound quality.

3.1 Level of Annotation and Analysis

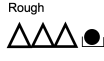



[28]In music, the level of annotation and analysis depends on the musical feature being analysed. Example 4 shows the different units of a song considered in this paper. Having a predefined level of analysis allows one to have a detailed representation of a piece while also drawing out salient musical points. In this paper, the tone quality features will be *annotated* at the level of the word. For words with multiple syllables, only the onset of the first syllable and only the termination of the last syllable will be annotated. Tone quality features will be *analysed* at the level of the phrase. This level of annotation and analysis is shown in Example 4. The level at which tone quality is annotated and analysed is shown by rectangles.

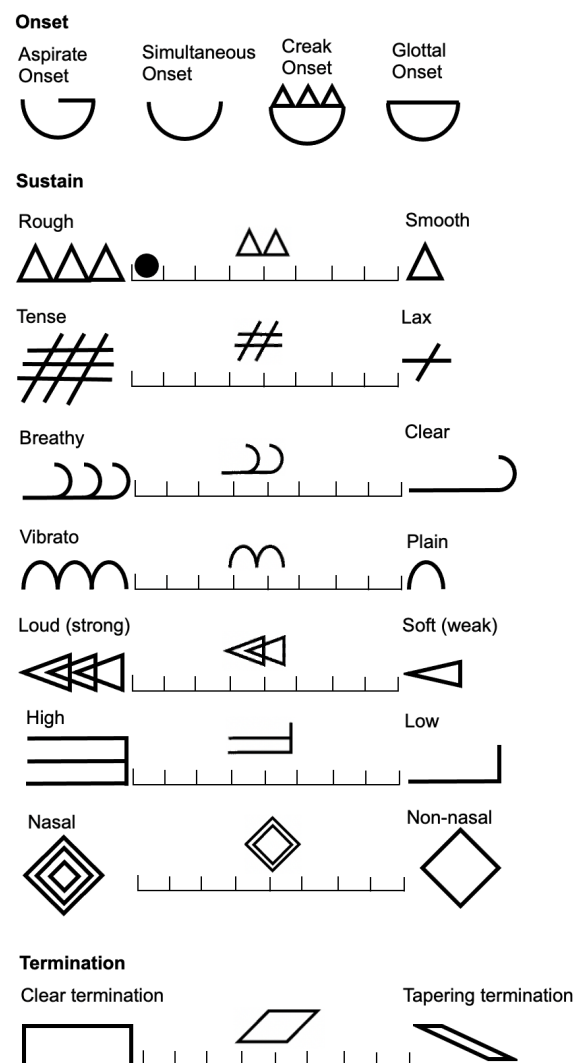
Level		Example	
Song		Whole song	
Section		Verse, chorus, bridge, etc	
Analysis	Musical Phrase	Defined by melodic contour, rhythm, harmony (i.e. not defined by lyrical content)	
Annotation	Word	One syllable word, e.g., song	Multi syllable word, e.g., singing
Note		Onset-sustain-termination	Onset-sustain-termination + Onset-sustain-termination

Example 4. The levels of a song considered in this paper. (The figure is the authors original work).

3.2 Tone Quality Features

[29]This section describes the tone quality features used for annotation and analysis. Example 5 shows the features explored in this section as well as their corresponding graphical representation. Each feature can be annotated either by using the scale or by using one of the discrete symbols available for each feature. For example, a tone quality may be assessed as being very rough (sitting at the extreme rough end of the scale), very smooth (at the extreme smooth end of the scale), or somewhere in between. One can either annotate the feature by

marking its position on the scale (e.g., to indicate very rough: , or utilising one of the three discrete symbols assigned to the feature (e.g., very rough: , very smooth: , moderate: ). Onsets are not graduated but are represented in terms of discrete symbols. This is because, as will be discussed in more detail in section 3.2.1 below, the four onset types given in Example 5 are considered sufficient for capturing the variation in the tone quality of an onset. The tone quality features can also be used to annotate a voice in real time through audio-video analyses. This is demonstrated in section 5 below.



Example 5. The tone quality features and their graphic representations. (The figure is the authors original work).

[30]The features presented below are not considered absolute, fixed points. Rather they are relative to the context of analysis – whether that be the conventions of a genre, a vocalist's unique tone quality, or variation within a single song. For this reason, before conducting analysis a baseline (or equipoise) should be identified (Author et al., 2021). The discussion of tone quality features within the analysis should be in reference to this equipoise.

[31]It is not only the configuration of a particular tone quality (e.g. how a single sound is graduated on each of the features in Example 5) but the amount of variability within a feature from one word/phrase to the next that impacts emotion perception. High variability of features has been found to be associated with certain emotive states. Variability refers to the amount of irregularity in a sound. Assessments of irregularity are not made in relation to an absolute reference point, but rather in relation to the surrounding use of that feature. This is identifiable at the level of the word but is most obvious at the level of the phrase. For example, a tone quality that consistently employed vibrato within a phrase would not be considered irregular, however one that varied from extreme plain to extreme vibrato between words would be considered irregular. Juslin and Laukka (2003) found that microstructural irregularity was associated with the basic emotions of anger, fear, and sadness (i.e. negative emotional states), while regularity was associated with happiness and tenderness (i.e. positive emotional states) (Juslin and Laukka 2003, Table 11). Going beyond the basic emotions, it is proposed in this paper that irregularity could also be associated with nervousness and anticipation, which are not necessarily negative but are states of uncertainty. Therefore, in this paper, I contend that high levels of variability signify uncertainty, while low levels signify certainty. In terms of its placement on the Affect Map, extreme high levels of variability are considered to be more unpleasant (valance), either awake or tired (arousal energy), and more tense (arousal tension).

3.2.1 Onset

[32]Onset forms the initial part of the note and is a highly characteristic musical feature (see, for example, Saldanha and Coroso, as cited in Erickson 1975, p. 61). I draw on three different classes of onset defined by Jo Estill (Mc Donald Kilmek, Obert, and Steinhauer 2005, 2 - 4): glottal onset, aspirate onset, and simultaneous onset. In addition, I add one more kind of onset to this list based on Heidemann (2016)— the creak onset. These four onsets are described in Table 1 below.

Table 1. A list of tone quality features that relate to onset. The first column gives the name of the feature, the second column provides a description of how the feature sounds/is produced, the third column provides implications for emotion perception. Column four describes where the feature in its extreme form may sit on the Affect Map.

Tone Quality Feature	Description	Perception	Position on Affect Map <i>1 Valance</i> <i>2 Arousal Energy</i> <i>3 Arousal Tension</i>
Glottal Onset	Glottal onsets are very percussive and are created by air “suddenly and forcefully escaping through a previously tightly closed glottis” (Heidemann, 2016, p. 6). Creak onsets are both percussive and rough and are produced by the vocal folds rapidly opening and closing as breath passes through.	Glottal and creak onsets may both be noisy (e.g., rough, breathy) and may therefore tend to be associated with negative emotional states (e.g., weariness, pain, sadness, fear) (Author, 2018, pp. 168 – 171). These onsets also tend to be very audible (it is difficult to produce a soft glottal onset, although creak onsets may have more dynamic variety). Therefore, in this paper I contend that glottal and, at times, creak onsets may also be associated with externalised negative emotions. That is, emotions which are typically expressed loudly and openly, either intentionally or involuntarily. For example, a scream of fear, or a yell of pain.	1 Either pleasant or unpleasant 2 More awake 3 More tense
Creak Onset			1 More unpleasant 2 Either awake or tired 3 Either tense or relaxed

Aspirate onsets	Aspirate onsets occur when breath passes through the vocal folds before they begin to vibrate (this may occur both while exhaling or inhaling). This creates a breathy effect which is noisy and unpitched, but not necessarily rough.	Aspirate onsets are like glottal/creak onsets in that they have the potential to be negative. However, they may signify a different kind of 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 of breathiness below).	<p>1 Either pleasant or unpleasant</p> <p>2 Either awake or tired</p> <p>3 Either tense or relaxed</p>
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.	Simultaneous onsets may be more likely to be associated with neutral and happy emotions. For example, 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.	<p>1 More pleasant</p> <p>2 More awake</p> <p>3 Either relaxed or tense</p>

3.2.2 *Sustain*

[33]Sustain forms the middle, and often the longest, part of the note. Because of its longer duration, there is ample opportunity for multiple tone quality features to become present. Sustain may also be highly varied as these features have the potential to undergo extreme changes within a short period of time. In developing these features I draw primarily theories of social semiotics of sound, specifically on the voice qualities outlined in van Leeuwen's discussion in *Speech, Music, Sound* (1999, pp. 129–141). Using my own findings (Author et al. 2019; Author 2018), these features have been expanded on such that they may be used to identify emotional valence in tone quality. It is not the goal here to provide a set of independent, individually measurable characteristics of a sound. Rather, all tone quality features are dialectically related – a sound is not only ever just breathy, it is always a combination of all available tone quality features. The analytical technique presented here, especially in the form of the real time annotations, is designed to accommodate for this by capturing the different configurations of tone qualities simultaneously. It is the examination of the different configurations of tone qualities within the voice, rather than measuring the degree of a single tone quality feature, which is the aim of this approach. These tone quality features are described in Table 2 below.

Table 2. A list of tone quality features that relate to sustain. The first column gives the name of the feature, the second column provides a description of how the feature sounds/is produced, the third column provides implications for emotion perception. Column four describes where the feature in its extreme form may sit on the Affect Map.

Tone Quality Feature	Description	Perception	Position on Affect Map <i>1 Valance</i> <i>2 Arousal Energy</i> <i>3 Arousal Tension</i>
Roughness: 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). Roughness is created by tensing the vocal folds and holding them tightly together (Heidemann, 2016, p. 6).	Noisy qualities may signify negative emotions (Author, 2018, pp. 168 – 171). For example, a scream is noisy, and a scream may be considered negative. Noisy sounds are related to roughness, which may result in roughness too being associated with negative emotions. However, there are different kinds of roughness. That is, in addition to the “tone of the voice itself”, other sounds that are present in a rough voice might range from extremely irregular aperiodic vibrations (like in screaming) to an equally rough but much more regular sound produced through consistent tension and air pressure (like in growling). Although it is true of all tone quality features, it is	Extreme roughness is: 1 More unpleasant 2 Either awake or tired 3 More tense

		especially the case that one must assess roughness against the equipoise of analysis (see paragraph 30 above).	
Breath: breathy/clear	<p>Breath can occur when “extraneous sound mixes in with the tone of the voice itself” (van Leeuwen, 1999, p. 133). It is produced when air leaks through an incompletely closed glottis (Heidemann, 2016, p. 5). When an aspirate 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).</p>	<p>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 as potentially expressing 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). I also suggest that breathiness in the voice may indicate vulnerability. For example, a sobbed utterance or the ragged breathy quality of a fearful voice.</p>	<p>Extreme Breath is:</p> <ol style="list-style-type: none"> 1 Either pleasant or unpleasant 2 Either awake or tired 3 Either tense or lax
Tension: tense/lax	<p>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</p>	<p>Tension may allow listeners to not only extrapolate information about a speaker’s physical state, but we may also gain a sense of their emotional state. For example, when we hear tension we may</p>	<p>Extreme tension is:</p> <ol style="list-style-type: none"> 1 Either pleasant or unpleasant

	results from tension not only is tense, it also means ‘tense’—and makes tense” (van Leeuwen, 1999, p. 131).	extrapolate that the speaker may be nervous, or in pain. Tension is produced alongside other tone quality features. For example, an extremely tense voice produced in the upper register may create the effect of a scream, while produced in the lower register with breath may sound more like a hiss. There are many situations in which tension may be present in the voice and the assessment of the emotional meaning of tensions is context dependant.	2 More awake 3 More tense
Vibrato: vibrato/plane	Vibrato is “a family of tonal effects in music” that is created by “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). In this way, vibrato and non-vibrato sounds may have the potential to evoke a range of responses. On the one hand, vibrato may signify love, tension, fear, and anticipation while non-vibrato may signify steadiness, an unmoving attitude, resolution, or acceptance (van Leeuwen, 1999, pp. 134 – 135).	Extreme vibrato is: 1 Either pleasant or unpleasant 2 More awake 3 Either relaxed or tense

Dynamic: loud/soft	<p>Dynamic here is related to performance intensity, which is “the loudness of the sound sources when ... performed in the recording studio process” (Moylan 2017, p. 139). This is distinct from the intensity (the actual volume) of the recording, which is the “energy transmitted by the sound wave across unit area per second”, the “duration and the frequency spectrum of the sound, and by the context in which the sound is heard” (Campbell and Greated, 2001, para. 1). In other words, Dynamic in this framework is concerned with the volume at which the sound was produced, rather than the volume at which it has been mixed into the recording. This is because this framework draws on the social semiotic experience of sound – the lived experience of speaking and listening to spoken voices.</p>	<p>Dynamic 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).</p>	<p>Extreme loudness is:</p> <ol style="list-style-type: none"> 1 Either pleasant or unpleasant 2 More awake 3 Either relaxed or tense
Range: high/low	<p>Range is connected to the changing location of sympathetic vibrations within the body (Heidemann, 2016, p. 8). The pitch of a sound is “determined by what the ear judges to be the most fundamental wave-</p>	<p>The use of high/low ranges may be associated with ideas of dominance and assertiveness. Van Leeuwen has found that men who mean to assert their dominance may speak in a higher range, while women who mean to do the same may speak in a lower range.</p>	<p>Extreme range (i.e., high pitch) is:</p> <ol style="list-style-type: none"> 1 Either pleasant or unpleasant

	frequency of the sound” (Haynes and Cooke, 2001, para. 1).	High/low singing also has an impact on tone quality. Falsetto may be used when men and women sing high. This falsetto results in a very different tone quality of the voice, when compared to singing in the lower range where intimate/soft sounds are more easily achieved. Singing in falsetto may also evoke ideas of effort, as to produce these sounds one must focus vibrations in the top of the head/sinuses (Heidemann, 2016, p. 8).	2 More awake 3 More tense
Nasality: nasal/non-nasal	Nasality is related to tension in that it is also produced by tensing the muscles, however nasality can also be produced through a cul-de-sac oscillation of air by allowing air to escape either through the nose or mouth only (van Leeuwen, 1999, p. 136).	Nasality in music has been shown to be a marker of sarcasm (Plazak, 2010, as cited in Huron, 2015, p. 190), and, especially for females, submission (Lomax, 1968 as cited in van Leeuwen, 1999, p. 137). Following on from this, and because of its close ties to tension, in this paper sounds which are very nasal are considered indicative of negative, high arousal emotions.	Extreme nasality is: 1 Either pleasant or unpleasant 2 More awake 3 More tense

3.2.3 *Termination*

[34] Termination forms the end of the note. The duration of a termination ranges from very short to very long. The table below describes these terminations and explores the emotional associations. This tone quality feature is described in Table 3 below.

Table 3. A list of tone quality features that relate to termination. The first column gives the name of the feature, the second column provides a description of how the feature sounds/is produced, the third column provides implications for emotion perception. Column four describes where the feature in its extreme form may sit on the Affect Map.

Tone Quality Feature	Description	Perception	Position on Affect Map <i>1 Valance</i> <i>2 Arousal Energy</i> <i>3 Arousal Tension</i>
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.	Tapered terminations allow for noise to become present in the sound, and therefore may be associated with negative emotions (Author, 2018, pp. 162 – 163). Clear terminations, on the other hand, may be associated with happier emotional states since there is less opportunity for extraneous sounds to become present. Take, for example, the ragged speech of a distraught individual (tapered termination) compared to the fast, clipped chatter of a happy child (clear termination).	Extremely clear termination is: 1 More pleasant 2 More awake 3 Either relaxed or tense

4 Analysing Text

[34]There are a range of approaches to text analysis in song. These range from explicit analysis of emotion words, to more nuanced sentiment analysis utilising theories such as Systemic Functional Linguistics (SFL) (see Author et al., 2021, for an overview of approaches and an example of SFL for lyric analysis). In this paper, I do not adopt a formal textual framework for analysis. I take this approach because text analysis sits independent of tone analysis. That is, one can take the tone analysis approach presented in this paper (tone quality features, Affect Map, and Cohesiveness) and apply whatever textual analysis they wish. Since the purpose of this paper is to present a system for annotating and analysing tone, and a tool for comparing this emotionality with text content, it is not necessary to mandate a form of text analysis. Indeed, the application of different text analysis systems (as has been explored in Author et al.'s SFL approach, 2021) to the analysis, and the analyses that result from this, provides an interesting point of future research. In this paper, I assess the emotionality of text based on my experience as a native English speaker, primarily drawing on words that explicitly convey emotion and that are emphasised in the musical phrase by the tone quality.

5 Application

[35]Having laid out the methodology in the above sections, the analytical approach will here be demonstrated through an analysis of the first verse and the first chorus of Kris Kristofferson's "Casey's Last Ride" taken from Kristofferson's 1970 album titled "Kristofferson". The structure of this song is built around two characters, Casey and The Woman. Example 6 shows the song timeline, characters, structure, and lyrics. The analysis of verse one (Casey) will be presented first, followed by the analysis of chorus one (The Woman).

[36]The equipoise of the song has been identified as line one of verse one (see Example 6, and hear this section in Audio-visual Example 1, 00:00 – 00:12). At this point, most of the tone quality features sit within the mid-point of the scale, onsets are mostly simultaneous. While some tone quality features tend towards the extremes of the scale (namely termination, breath, and vibrato), there is little variability in the features – that is, there is a consistency of tone quality which makes this a good example of tone quality equipoise.

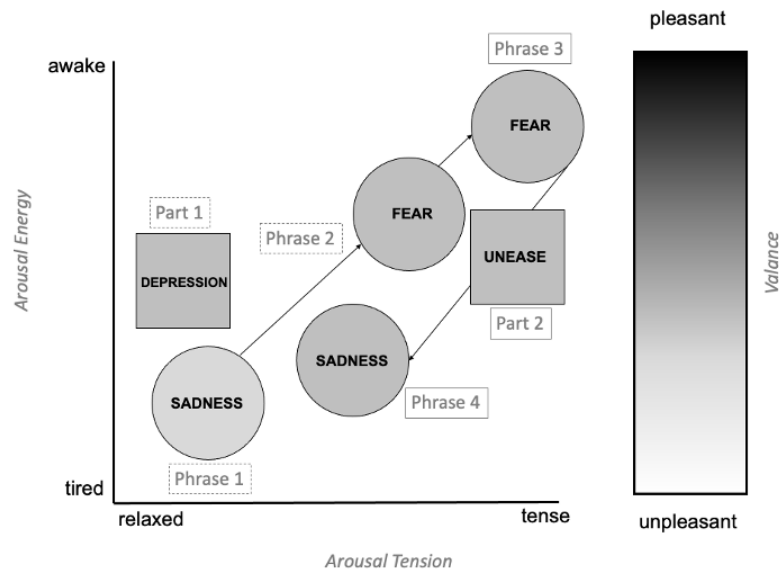
Time	Structure	Character	Line	Phrase	Part	Lyrics
0:00	Intro					
0:07	Verse One	Casey	1	1	1	Casey joins the hollow sound of silent people walking down
			2			The stairway to the subway in the shadows down below;
			3	2		Following their footsteps through the neon-darkened corridors
			4			Of silent desperation, never speakin' to a soul.
			5	3	2	The poison air he's breathin' has the dirty smell of dying
			6			'Cause it's never seen the sunshine and it's never felt the rain.
			7	4		But Casey minds the arrows and ignores the fatal echoes
			8			Of the clickin' of the turnstiles and the rattle of his chains.
0:58	Interlude					
1:04	Chorus One	The woman	1	1		Oh! she said, Casey it's been so long since I've seen you!
			2	2		Here she said, just a kiss to make a body smile!
			3	3		See she said, I've put on new stockings just to please you!
			4	4		Lord! she said, Casey can you only stay a while?
1:39	Interlude					
1:46	Verse Two	Casey	1	1	1	Casey leaves the under-ground and stops inside the golden crown
			2			For something wet to wipe away the chill that's on his bone.
			3	2		Seeing his reflection in the lives of all the lonely men
			4			Who reach for any thing they can to keep from goin' home.
			5	3	2	Standin' in the corner Casey drinks his pint of bitter
			6			Never glancing in the mirror at the people passing by
			7	4		Then he stumbles as he's leaving and he wonders if the reason
			8			Is the beer that's in his belly, or the tear that's in his eye.
2:36	Interlude					
2:43	Chorus Two	The woman	1	1		Oh! she said, I suppose you seldom think about me,
			2	2		Now she said, now that you've a family of your own;
			3	2		Still she said, it's so blessed good to feel your body!
			4	4		Lord! she said Casey it's a shame to be alone!
3:17	Outro					
3:38						

Example 6. Timeline of the song “Casey’s Last Ride” by Kris Kristofferson.

Verse 1

[37] Casey is introduced in verse one. The lyrics progress from a sense of depression to a sense of unease. The lyrics in phrase one are more relaxed, tired, and unpleasant (see Example 7) as they outline Casey's complicit descent into the subway. This is shown in Example 7 where the Affect Map is used to represent the emotions in this verse. Phrase, part and lyrics are given below the Affect Map. This configuration of arousal and valence is generally considered consistent with sadness (see Section 2.1). Phrase two moves to a more tense, awake, and moderately unpleasant emotionality (see Example 7) as Casey describes the isolation and silent desperation of his descent (see Example 7). Such a configuration might suggest a moderate fear (see Section 2.1). The mixture of sadness and fear present in part one creates a sense of depression – Casey understands the desperateness, yet helplessness, of his situation.

[38] Phrase three conveys a lyrical message that is mostly tense, awake, and moderately unpleasant (see Example 7), which is typically associated with fear (see Section 2.1). This is evoked by the use of highly emotive words which relate to death such as “poison” and “dying”, as well as words that signify deprivation such as “never felt the rain”. Phrase four returns to sadness, but this time it is more tense, awake, and unpleasant than in phrase one (see Example 7). The movement of fear to more intense sadness creates a mood of unease in part two (Example 7). The lyrics point to Casey knowing of the danger he is in, but suggest an inability to do anything to change the situation, as emphasised by the phrase “clicking of the turnstile” and “rattle of his chains” which suggest that Casey is continuing in his established patterns.



Phrase	Part	Lyric
1	1	Casey joins the hollow sound of silent people walking down The stairway to the subway in the shadows down below;
2		Following their footsteps through the neon-darkened corridors Of silent desperation, never speakin' to a soul.
3	2	The poison air he's breathin' has the dirty smell of dying 'Cause it's never seen the sunshine and it's never felt the rain.
4		But Casey minds the arrows and ignores the fatal echoes Of the clickin' of the turnstiles and the rattle of his chains.

Example 7. An assessment of the emotion present in the lyrics of verse one, “Casey’s Last Ride”.

[39]The tone quality of this verse has been annotated in real time using the tone quality features discussed above. See Audio-visual Example 1 for this real time annotation. Example 6 shows the lyrics of this verse. As with the assessment of emotion in lyrics above, an analysis of tone quality using Audio-visual Example 1 reveals that verse one can be split into two equal parts consisting of lines 1 – 4 and 5 – 8 (see Example 6). This delineation can be observed through each of the categories of tone quality features: onset, sustain, and termination.

[40]Onsets in part one are mostly aspirate and simultaneous. By contrast, the onsets in part two are much more varied. In part two, glottal and creak onsets play a greater role than part one, especially around line 5 – 7. This increased use of glottal and creak onsets is shown in

Example 8. Note that in verse one aspirate onsets do occur but only on words which would typically have an aspirate onset anyway (e.g. words beginning with “s” and “f”). For this reason, aspirate onsets are not annotated nor analysed for their emotionality as they are considered to be by-products of the pronunciation of lyrics. This is also the case for simultaneous onsets which are not annotated or analysed in Example 8 since they are considered to be most neutral and also by-products of lyric pronunciation.

Line	Phrase	Part	Lyrics
1	1	1	Casey joins the hollow sound of silent people walking down
2			The stairway to the subway in the shadows down below;
3	2		Following their footsteps through the neon-darkened corridors
4			Of silent desperation, never speakin' to a soul.
5	3	2	The poison air he's breathin' has the dirty smell of dying
6			'Cause it's never seen the sunshine and it's never felt the rain .
7	4		But Casey minds the arrows and ignores the fatal echoes
8			Of the clickin' of the turnstiles and the rattle of his chains.

Key:

Aspirate Onset	Simultaneous Onset	Creak Onset	Glottal Onset

Example 8. Glottal and creak onsets in the first verse of Casey's Last Ride.

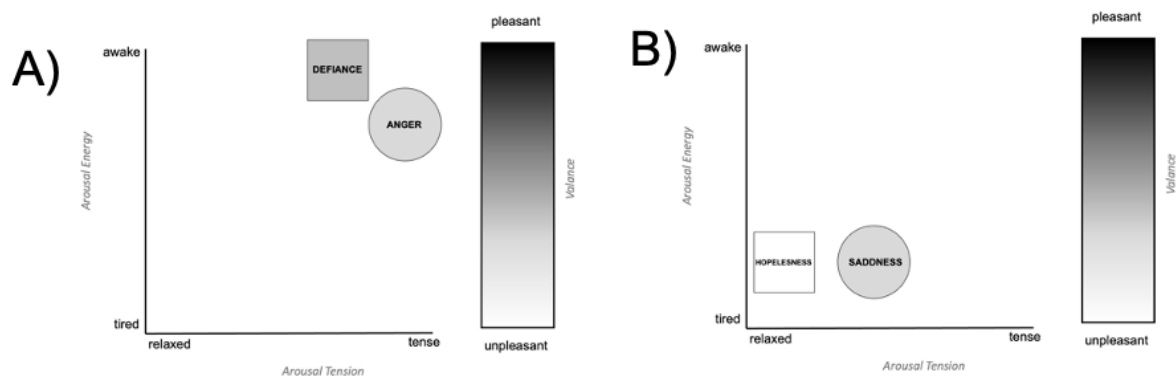
[41]Accompanying this change in onset, there is also a change in lyrical content in part two.

Whereas part one alludes to a monotonous, isolating experience of life, part two refers directly to death – either metaphorical or literal. It is on words that make references to death that glottal and creak onsets can be heard.

[42]Phrase 3 is the first time a glottal onset is heard in part two. While the word it occurs on “breathin’”, is not a direct reference to death itself, the words which precede it, “The poison air”, cast the word “breathin’” in a negative light. It is the act of breathing this poison air that signifies death. Phrase 3 contains four more glottal onsets on the words “dirty”, “dying”,

“never”, and “rain” (Example 8). Each time, these onsets draw out the lyrical message of death and stagnation. Glottal onsets bookend the phrase “dirty smell of dying”, drawing out the negative message of this phrase. Similarly, glottal onsets also bookend “never felt the rain”, which highlights the message of stagnation. In phrase 4, the glottal onsets on “never” and “ignores” serve to further draw attention to the deadly stagnation of the poison air by bookending the phrase “ignores the fatal echoes”. Before examining the second half of phrase 4, let’s first pause to consider the impact of these glottal onsets on emotional perception.

[43]The glottal onsets in part two signify an emotional state that is tense, mostly awake, and moderately unpleasant (Example 9). This may be perceived as indicative of anger. The regular use and placement of glottal onsets in part two (compared with part one) creates an overall mood of defiance. Casey reacts angrily to the lyrical message of death in part two, he is defiant, and, in this defiance, there is hope that the situation may yet change. However, the second half of phrase 4 colours this message with a different emotion and mood.

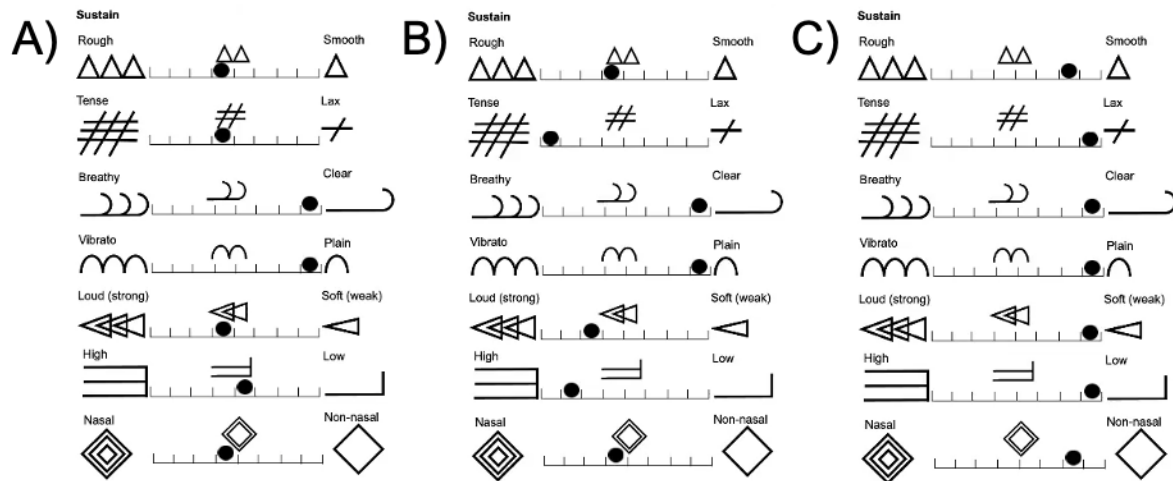


Example 9. Assessing the onsets in part two of verse one against the Affect Map. A) shows the assessment for lines 5 – 7, B) shows the assessment for line 8.

[44]The second half of phrase 4 has both a glottal and a creak onset (Example 8). The glottal onset on “clickin” appears to perpetuate this sense of defiance. However, its placement is different to the other phrases (occurring very early in the line), and it falls on a word which

does not appear to convey a message of death and stagnation. The proceeding creek onset on “rattle” may shed some light on this. This creak onset signifies a mostly relaxed, tired, unpleasant emotive state (see section 3.2.1). This may be perceived as indicative of sadness. The use of a single creak onset and the resulting sense of sadness creates an overall mood of hopelessness (Example 9). This sense of hopelessness is further heightened when we consider the previous glottal onset. Compared to the rest of the verse, line 8 has high variability of onsets both in kind (creak + glottal) and placement (glottal at the start of the line). As discussed in section 3.2, high variability is associated with negative emotional states. In this way, Casey’s defiance is betrayed by the onsets in line 8. Has Casey’s resolve been rattled? Does the knowledge of his impending fate penetrate his defiance, just as the irregularity of line 8 penetrates through the texture of the song?

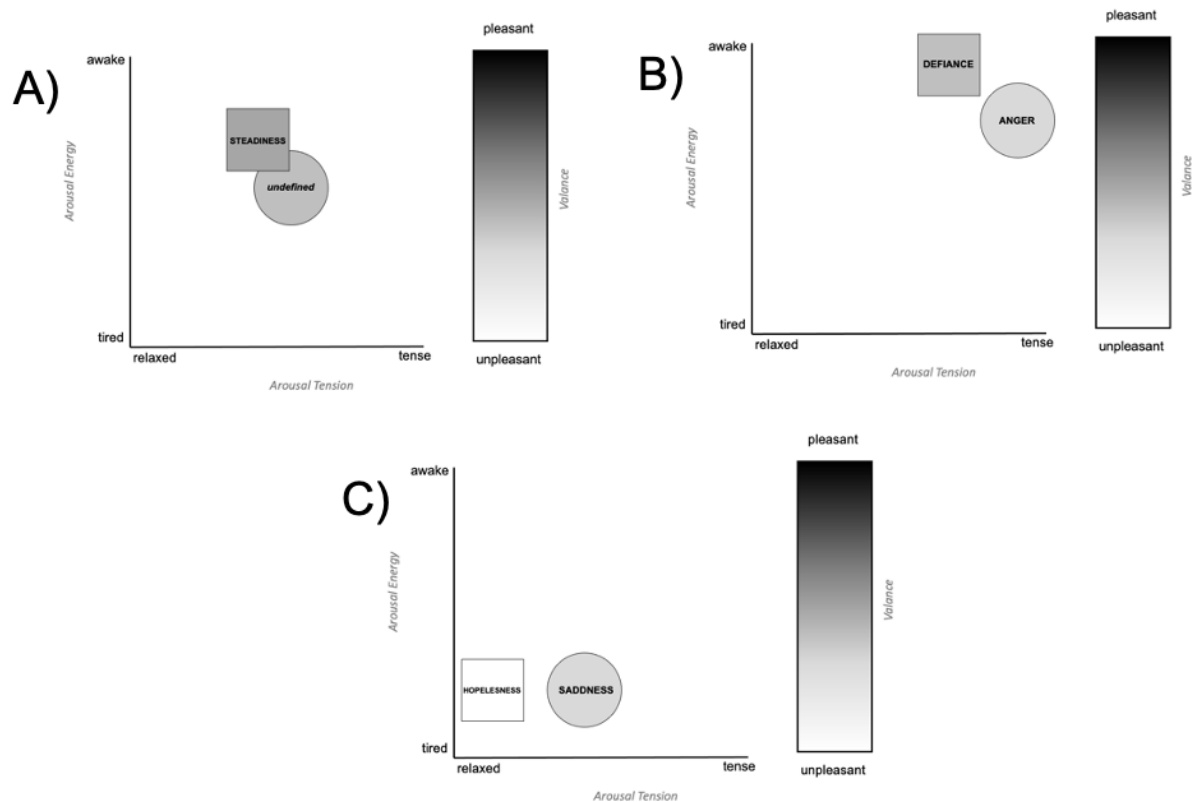
[45]The emotional message conveyed through onset is reinforced through the sustain. Part one of the verse is quite steady across the tone quality features of sustain (Example 10). The tone is consistently clear and plain. The ends of each line tend to become softer and lower, and tension and roughness oscillate between low and medium. There is moderate nasality which, as discussed in section 3.2.2, is indicative of negative emotional states. However, taken with the steady presentation of the other tone quality features, overall sustain in part one is mostly relaxed, awake and pleasant (Example 11). This is not indicative of any particular emotive state, but it does create a mood of stability (Example 11).



Example 10. Annotation of tone quality features which are exemplars of sustain in verse one.

A) shows the assessment for part one, B) shows the assessment for part two lines 5 – 7, C) shows the assessment for part two, very end of line 8.

[46]The sustain in the second half of the verse, like the onsets, paints a different picture (Example 10). Suddenly, tension becomes present in phrases three and four. The level of roughness also varies quite suddenly and obviously in the second half of phrase four. A greater use of the upper range is apparent here too, especially in line six. Nasality remains relatively consistent, however the potential negative connotations of moderate nasality in part one are now realised in part two with increasing variability in the other features. The presentation of the features at their more extreme ends, and the increasing variability within part two suggest an emotive state that is more tense, awake, and unpleasant (Example 11). This may be indicative of anger (especially given the plain, clear delivery), and may contribute to the mood of defiance suggested by the onsets. Like onset too, the final line becomes gradually more lax, low, non-nasal and soft (Example 10), which contributes to the perception that Casey's emotional state has become mostly relaxed, tired, and unpleasant (Example 11). This is indicative of sadness and hopelessness (Example 11).

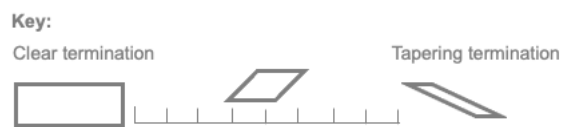


Example 11. Assessing the sustain of verse one against the Affect Map. A) shows the assessment for part one, B) shows the assessment for part two lines 5 – 7, C) shows the assessment for part two line 8.

[47]Termination is the least varied of the tone quality features. Terminations remain mostly strong throughout, with only three instances of weaker terminations in phrases 2 and 4 (Example 12). These weaker terminations occur at important structural points within the verse, on lines 4 and 8, which contributes to the two-part structure of verse one (Example 12). The tapering termination at the end of the verse is particularly interesting. As discussed in section 3.2.3, tapering terminations can signify negative emotions which may be more relaxed, tired and unpleasant. The tapering termination at the end of phrase 4, then, is in line with the mood and emotion conveyed through onset and sustain (Examples 9 and 11). This

reinforces the emotional message conveyed through tone quality – in verse one Casey has been through a journey of control, defiance, and finally hopelessness.

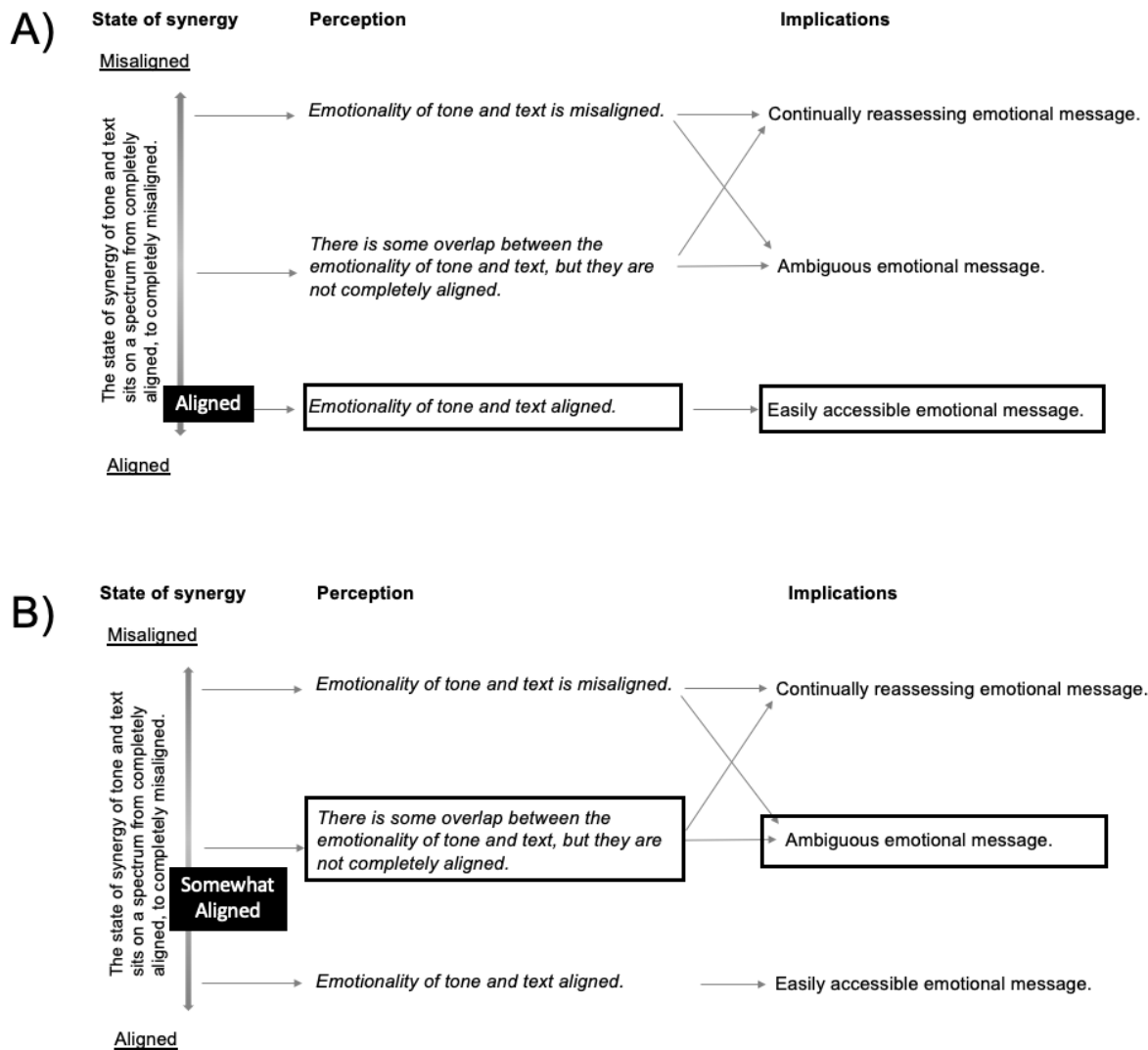
Line	Phrase	Part	Lyrics
1	1	1	Casey joins the hollow sound of silent people walking down
2			The stairway to the subway in the shadows down below;
3	2		Following their footsteps through the neon-darkened corridors
4			Of silent desperation , never speakin' to a soul .
5	3	2	The poison air he's breathin' has the dirty smell of dying
6			'Cause it's never seen the sunshine and it's never felt the rain.
7	4		But Casey minds the arrows and ignores the fatal echoes
8			Of the clickin' of the turnstiles and the rattle of his chains .



Example 12. Terminations in the first verse of Casey's Last Ride.

In sum

[48]The overall message delivered in the vocal line in part one of verse one is mostly aligned (Example 13). Tone quality features in this section do not tend towards the extreme of any emotion, but rather create a sense of steadiness (Example 11). This moderate tone does not conflict with the text through which a general sense of depression is created (Example 7). Thus, the tone/text relationship in the first half of the verse creates an affirming message for the listener, it is easy to access and understand the emotionality of the vocal line (Example 13).



Example 13. Using cohesiveness to assess the state of synergy of emotion conveyed through tone and text in verse one of “Casey’s Last Ride”. A) shows the state of synergy for part one,

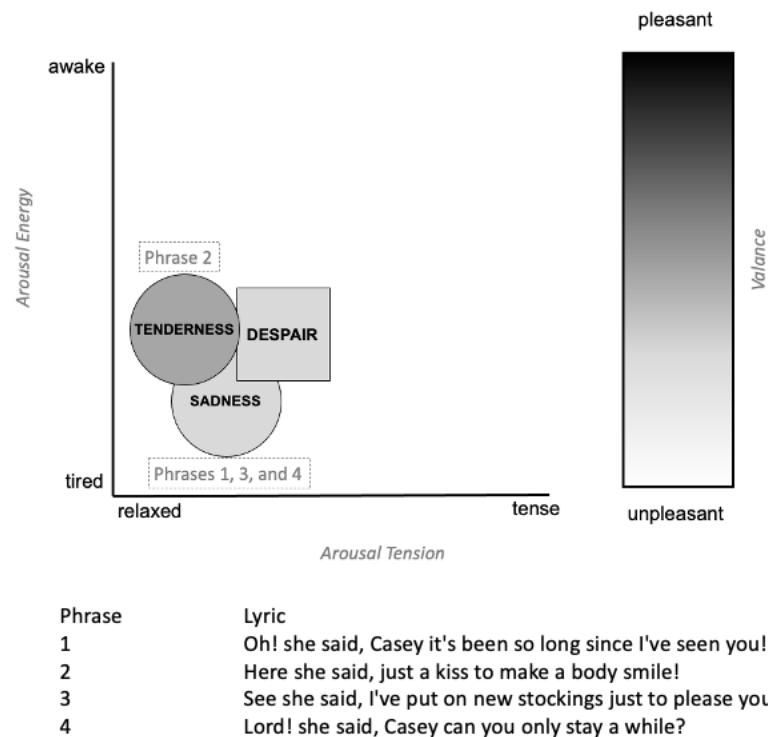
B) shows the state of synergy for part two.

[49]The second half of the verse, on the other hand, presents a slightly different tone/text relationship. The lyrics in the second half create a sense of unease (Example 7). The tone quality, however, moves from a sense of defiance in lines 5 – 7, to a sense of hopelessness in line 8 (Examples 9 and 11). While ending the verse with a tone quality which suggests hopelessness doe create a sense of unease, in general the tone and text remain only somewhat

aligned (Example 13). In general, the listener must reflect on tone and text to understand the implications of the unease and defiance in lines 5 – 7 (has Casey broken free of the depression in part one?), as well as assess the meaning of the suddenly aligned message at the end of unease and hopelessness (did Casey give in to the depression which surrounded him in part one?).

Chorus 1

[50]The Woman is introduced in chorus one. Example 6 shows the lyrics for this section. The lyrics are pleading as The Woman tries to convince Casey to stay with her. The lyrics of phrases 1, 3, and 4 are mostly relaxed, moderately tired and mostly unpleasant. This is shown in Example 14 where the Affect Map is used to represent the emotions in this chorus. Phrase, part and lyrics are given below the Affect Map. It is in these phrases that The Woman makes her pleas to Casey – she has missed him and now that he is here, can't he stay just a little longer? This plea is not explicit, but rather it is conveyed implicitly through the word phrases. This implicitness contributes to the more relaxed, tired emotionality of these phrases – the emotion is subdued and under the surface. The lyrics are also more unpleasant as they suggest that the relationship never used to be this distant (Example 14). This configuration of arousal and valence is generally consistent with sadness (see Section 2.1).



Example 14. An assessment of the emotion present in the lyrics of chorus one, “Casey’s Last Ride”.





[51]The lyrics of phrase 2, however, convey a different message. Here, the lyrics are also relaxed and tired, conveying the emotional message implicitly (Example 14). However, this phrase is more pleasant due to the use of words such as kiss, and smile, as well as the implication of closeness to Casey (Example 14). This configuration of valence and arousal might be considered consistent with tenderness – The Woman still loves Casey even if he is making her sad (Example 14). This mixture of sadness and tenderness creates an overall sense of despair.

[52]The tone quality of this chorus has been annotated in real time using the tone quality features discussed above. See Audio-visual Example 2 for this real time annotation. Example 6 shows the lyrics of this verse. As can be seen in Audio-visual Example 2, aspirate and simultaneous onsets are used exclusively throughout this chorus except for phrase 4. On this

phrase, glottal and creak onsets are used (Example 15). Similar to the analysis of onset above, aspirate and simultaneous onsets are not annotated in Example 15 since they tend to be by-products of the pronunciation of lyrics. The use of such onsets does not suggest any highly charged emotional state, rather it is quite moderately tense, awake, and pleasant, creating a mood of stability (Example 16). Against this aural backdrop, the use of glottal and creak onsets in phrase 4 is quite salient.

Line	Phrase	Lyric
1	1	Oh! she said, Casey it's been so long since I've seen you!
2	2	Here she said, just a kiss to make a body smile!
3	3	See she said, I've put on new stockings just to please you!
4	4	Lord! she said, Casey can you only stay a while?

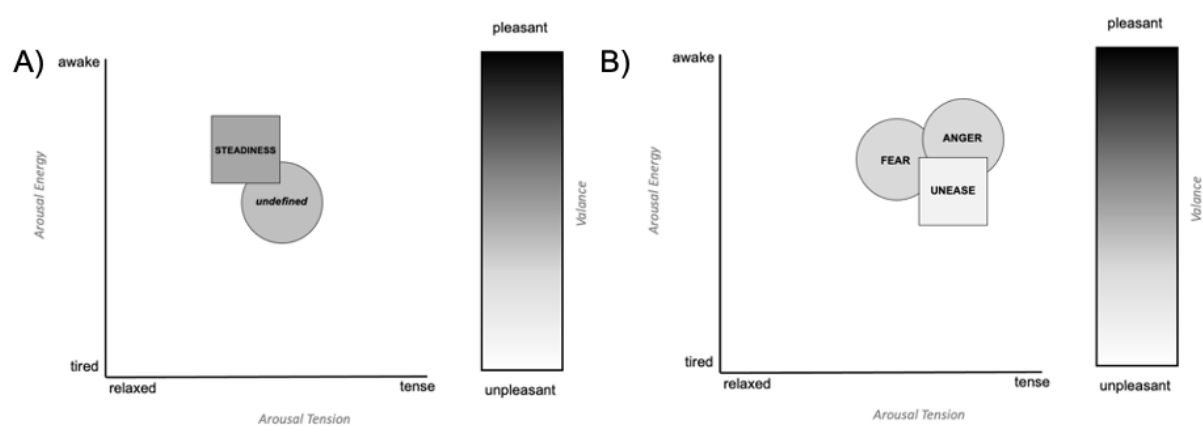
Key:

Aspirate Onset	Simultaneous Onset	Creak Onset	Glottal Onset
			

Example 15. Glottal and creak onsets in the first verse of Casey's Last Ride.

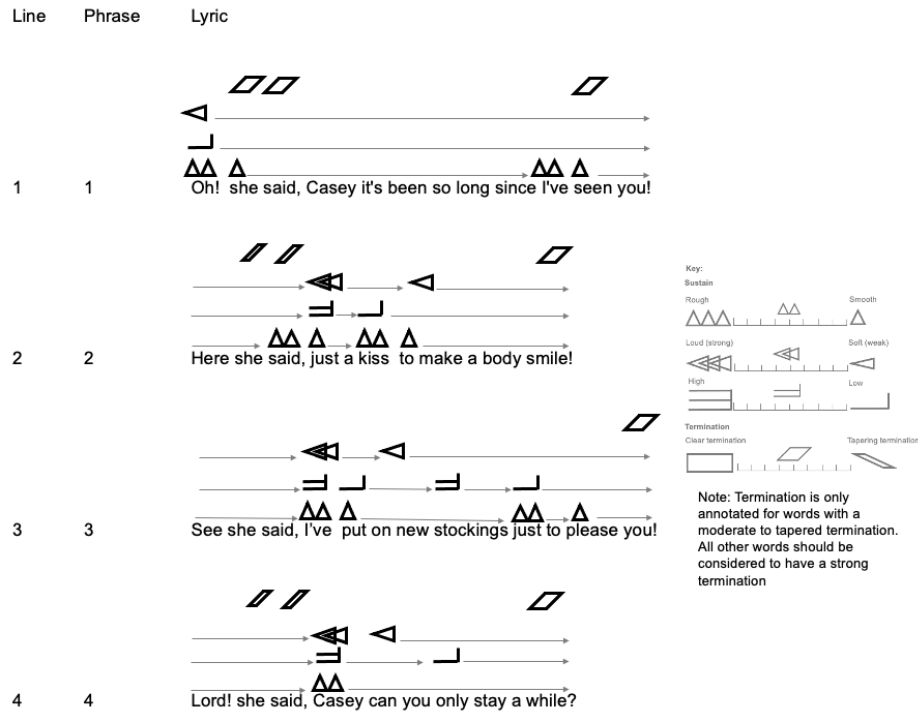
[53] This change of onset is also accompanied by a change in lyrical message. In the first three phrases The Woman is speaking in statements. In phrase four, The Woman poses a question for the first time. The question begins almost as a demand with the strong, assertive glottal onset on "Casey", but ends as a beg with the creak onset on the word "only". This sudden questioning and variation in onset increases the salience of phrase 4. The glottal onset may signify a mostly tense, awake, and moderately unpleasant emotional state (Example 16). Such a configuration may suggest anger – The Woman is aggressive in her demand for Casey to stay. The creak onset, however, subverts this emotional message. The slightly more relaxed and tired arousal of this onset is more consistent with fear – it is at this point that The Woman tempers her message with the use of the word "only" (Example 16). The variability

in onsets, although small, is salient and heightens the sense of the negative mood created in this chorus. Taken together, onsets in chorus one create a mood of unease (Example 16).



Example 16. Assessing the onsets in chorus one against the Affect Map. A) shows the assessment phrases 1, 2, and 3, B) shows the assessment for phrase 4.

[54]The sustain and termination reflect this message too. The entire chorus is almost always plain, breathy, and non-nasal, but variation can be seen in the other vocal features (see Audio-visual Example 2). Variations in range, dynamic, and roughness begin to be heard from phrase 2. Here, the word “just” is stronger and higher than the vocal quality in the previous line (Example 17). However, this is short lived, with a return to the low, soft vocal quality on the following words, “a kiss” (Example 17). But the delivery of “a kiss” is not smooth as the preceding vocal quality has been. Instead, it is rough. While this is not the first-time roughness has become present in this otherwise soft, low voice, it is the first time it has been preceded by any other variation in sustain. Indeed, the words “just a kiss” are accentuated by the unusually weak termination on the words “she said”.



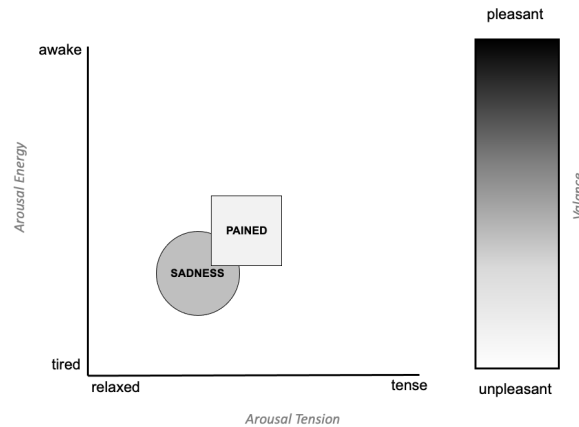
Example 17. Annotation of roughness, dynamic, range, and termination for chorus 1 of

“Casey’s Last Ride”.

[55] Phrases 3 and 4 also exhibit variations in sustain and termination. Phrase 3 begins with the higher delivery of the personal pronoun “I” however as the phrase progresses the sustain becomes lower and softer again (Example 17). The delivery of the final words “to please you” sees a return to the low, weak sustain. This is underscored by the termination with which the final word is delivered. Phrase four begins much like the previous, the weaker termination on the words “she said” followed by the stronger, higher delivery of “Casey” suggests an assertiveness (Example 17). However, this is short lived as the final words “only stay a while” are delivered with a consistent roughness, persisting for the first time almost the entire length of the phrase.

[56] Throughout the chorus, sustain and termination create an emotional message that is mostly relaxed, tired, and unpleasant (Example 18). This configuration is consistent with

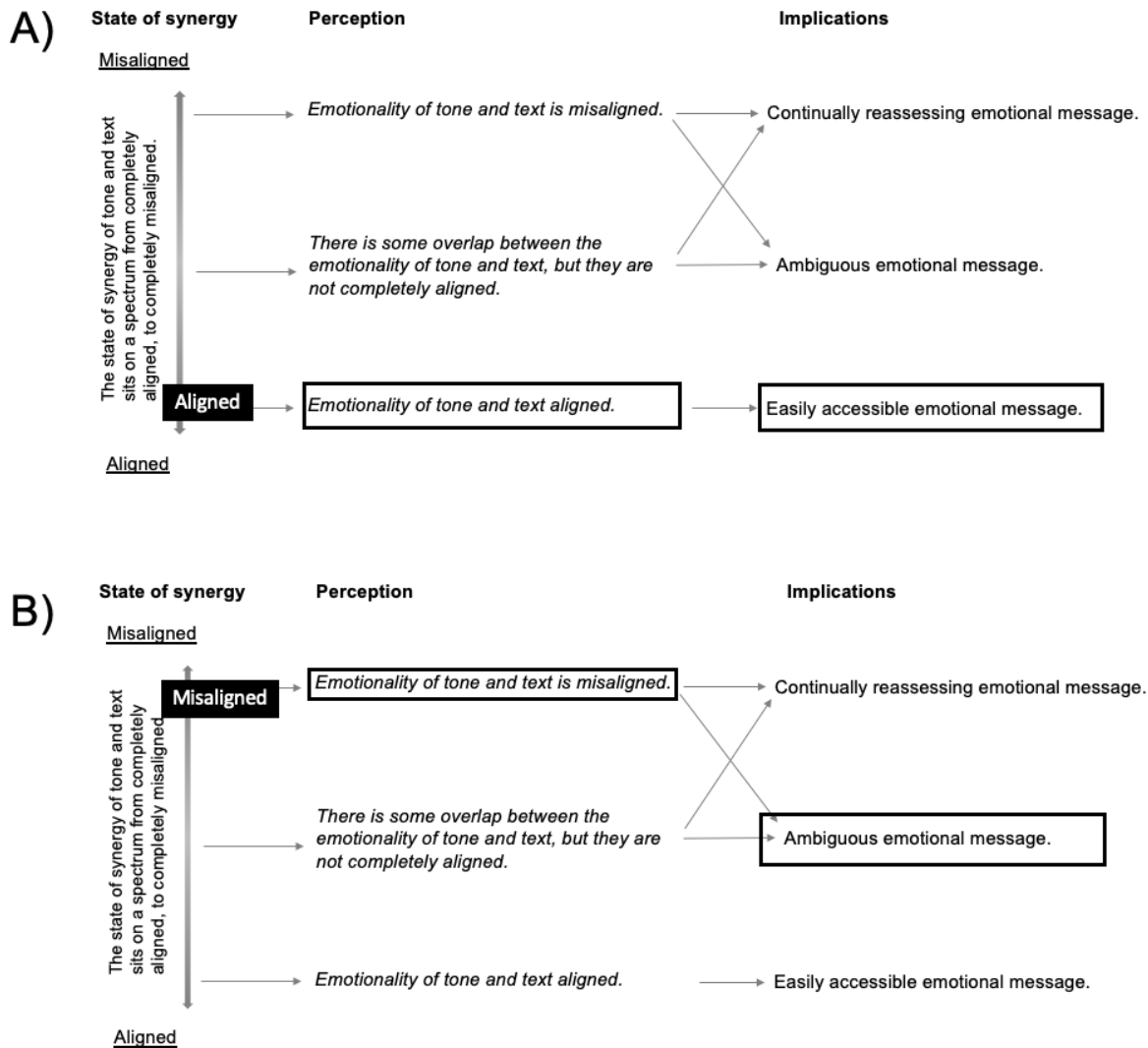
sadness (Example 18). The combination of consistency in breath, nasality and vibrato, combined with relatively high variability in roughness, dynamic, range, and termination, heighten the negativity of this emotion, creating a mood of pain and helplessness.



Example 18. Assessing the sustain and termination chorus one against the Affect Map.

In sum

[57]The overall message delivered in the vocal line in chorus one is mostly aligned (Example 19). Tone quality features suggest low arousal and negative valence, creating a mood of pain (Example 16 and 18). On the whole this does not conflict with the text through which a general sense of despair is evident (Example 14). Thus, the tone/text relationship in chorus one creates an affirming message for the listener, it is easy to access and understand the emotionality of the vocal line (Example 19).



Example 19. Using cohesiveness to assess the state of synergy of emotion conveyed through tone and text in chorus one of “Casey’s Last Ride”. A) shows the state of synergy for the chorus overall, B) shows the state of synergy for the tone/text relationship on the words “just a kiss” and “just to please you”.

[58]However, there are instances where vocal quality and lyrics are misaligned (Example 19). In particular, the lyrics “just a kiss” and “just to please you” which might be considered positive emotional messages are delivered with a negative tone. While brief, these small instances of misalignment are salient. This creates a more ambiguous emotional message, causing the listener to reassess the tone/text relationship to ascertain the overall message

(Example 19). The result is that at times the emotional message is not always obvious.

Rather, it unfolds over time, making the listener work to interpret the message on a phrase by phrase, word by word basis. Ultimately, this serves to heighten the emotional message, the listener, uncertain of what is happening, struggles through the verse just as The Woman struggles through her final encounter with Casey.

Contrasting verses and choruses

[59]There is a noticeable contrast between the tone quality of verse one, which relates to Casey, and chorus one, which relates to The Woman. Overall, verse one is abrasive, while chorus one is anguished. There are several explanations for this contrast.

1. Casey always acts, while The Woman always speaks. Casey's actions are central to his story. The listener observes Casey moving through a hard, rough world, a reality that is manifested in the tone quality. The Woman, on the other hand, exists in bittersweet memory which is signified by the soft and fervent tone quality.
2. Casey is the present, The Woman is the past. This dichotomy is established from the first few lines – Casey is following and seeing. This present is uncertain and dangerous, and this is reflected in the generally high arousal, unpleasant tone quality (and further underscored by variations in the two parts of the verse). The Woman, on the other hand, is the past – she is not saying, but has said. The past is fixed, and so too then are the negative events of the past (The Woman not being able to make Casey stay). This sense of helplessness is reflected in the tone quality and the emotional vulnerability is heightened by changing cohesiveness in the text/tone relationship between the chorus in general and key phrases in particular.
3. Casey is death, while The Woman is life. Literal and metaphorical death is a salient lyrical theme in verse one, as the listener observes Casey's isolating and lonely

experience. Chorus one, however, is shaped by The Woman's connection to Casey (seeing, speaking, kissing). The Woman is life immortalised – her place in memory means that she cannot be touched by the harsh, physical realities of the present world, including death. She anchors Casey to the world of the living, calling for Casey to stay, if only for a while.

[60]In both tone and text, Casey and The Woman sit at opposite ends of the spectrum. This divide between the characters heightens their individual messages. Once intimately connected, the characters are now separated from one another, and their narratives express this loneliness in contrasting ways. This contrast both in text and in tone drives the song forward.

Conclusion

[61]The goal of this paper was to present a new analytical approach for the sung voice. This was achieved by presenting a new framework for analysing tone quality through the tone quality features and associated tools (the Affect Map and Cohesiveness). This paper also offers a method of considering the emotionality of text alongside that of tone. As discussed above, the analytical approach proposed in this paper is not intended to be prescriptive. Instead, the goal is to provide a consistent framework to annotate, analyse and describe tone quality and its relationship with text. Others, through applying the same framework, may arrive at different conclusions about emotion and mood in tone quality and text. This is not uncommon in music analysis. Indeed debating different conclusions drawn from the application of the same framework is a regular occurrence – for example, different interpretations may be drawn from the application of Schenkerian Analysis to the same piece. Many of the tone quality features here, specifically those used to describe sustain, are adopted from the social semiotics of sound proposed by van Leeuwen (1999). This is based primarily on the acoustic experience of sound. In future work, how the tone quality features can be

extended to account for technologically mediated sound experience (e.g. in relation to performance v sound intensity for Dynamic) is an important avenue for research. This must continue to be balanced, however, with useability.

The main implications for this approach relate to the annotation and analysis of the voice in popular vocal songs, something for which there have been few frameworks for discussing in the past. Additionally, this paper draws on psychology and social semiotics to offer a systematic method of assessing emotionality of tones and for integrating this into analysis. In this way, this paper has presented a framework on which future research can build in a variety of ways, including through musicology, multimodality, semiology, and psychology.

Works Cited

Author et al, 2019.

Author et al., 2021

Author, 2018.

Campbell, Murray, and Clive Greated. 2001. Loudness. In *Grove Music Online*.

Carter, Tim. 2001. "Word-Painting." In Oxford Music Online. Oxford University Press.

<https://doi.org/10.1093/gmo/9781561592630.article.30568>.

Eerola, T., and J.K. Vuoskoski. 2011. "A comparison of the discrete and dimensional models of emotion in music." *Psychology of Music* 39 (1): 18-49.

<https://doi.org/10.1177/0305735610362821>.

Eerola, Tuomas, and Jonna K. Vuoskoski. 2013. "A Review of Music and Emotion Studies: Approaches, Emotion Models, and Stimuli." *Music Perception* 30 (3): 307-340.

<https://doi.org/10.1525/MP.2012.30.3.307>.

Erickson, R 1975. *Sound Structure in Music*. Berkeley, Los Angeles, London.: University of California Press.

Evans, Paul, and Emery Schubert. 2008. "Relationships between expressed and felt emotions in music." *Musicae Scientiae* 12 (1): 75-99.

<https://doi.org/10.1177/102986490801200105>.

Fairclough, N. 2001. "The Discourse of New Labour: Critical Discourse Analysis." In *Discourse as Data: A Guide for Analysis*, edited by M. Wetherell, S. Taylor and S.J. Yates, 229-266. London: Sage Publishing.

Frith, Simon. 1998. *Performing Rites: On the Value of Popular Music* Cambridge, Massachusetts: Harvard University Press.

Haynes, Bruce, and Peter Cooke. 2001. Pitch. In *Grove Music Online*.

- Heidemann, K. 2016. "A System for Describing Vocal Timbre in Popular Song." *Journal for the Society for Music Theory* 22 (1).
<http://www.mtosmt.org/issues/mto.16.22.1/mto.16.22.1.heidemann.html>.
- Hunter, Patrick G., and E. Glenn Schellenberg. 2010. "Music and Emotion." In *Music Perception*, edited by Mari Riess Jones, Richard R. Fay, and Arthur N. Popper, 36:129–64. Springer Handbook of Auditory Research. New York, NY: Springer New York. https://doi.org/10.1007/978-1-4419-6114-3_5.
- Huron, D. 2015. "The Other Semiotic Legacy of Charles Sanders Peirce: Ethology and Music-Related Emotion." In *Music, Analysis, Experience: New Perspectives in Musical Semiotics*, edited by C. Maeder and M Reybrouck, 185 - 208. Belgium: Leuven University Press.
- Juslin, Patrik N., and Petri Laukka. 2003. "Communication of Emotions in Vocal Expression and Music Performance: Different Channels, Same Code?" *Psychological Bulletin* 129 (5): 770–814. <https://doi.org/10.1037/0033-2909.129.5.770>.
- Kroeger, Karl. 1988. "Word Painting in the Music of William Billings." *American Music* 6 (1): 41. <https://doi.org/10.2307/3448345>.
- Lacasse, S. 2010. "Slave to the supradiegetic rhythm: A microrhythmic analysis of creaky voice in Sia's 'Breathe Me'." In *Musical Rhythm in the Age of Digital Reproduction*, edited by A. Danielsen, 141 - 158. London, UK: Routledge.
- Lavan, N., A.M. Burton, S.K. Scott, and C. McGettigan. 2018. "Flexible voices: Identity perception from variable vocal signals." *Psychonomic Bulletin & Review*.
<https://doi.org/https://doi.org/10.3758/s13423-018-1497-7>.
- Laver, J. 1980. *The Phonic Description of Voice Quality*. Cambridge Cambridge University Press.

- Lomax, A. 1968. *Folk song style and culture*. Washington, DC: American Association for the Advancement of Science.
- McDonald Kilmek, Mary , Kerrie Obert, and Kimberly Steinhauer. 2005. *Estill Voice raining, Level Two: Example Combinatiosn for Six Voice Qualities* Pittsburgh: Estill Voice Training Systems International, LLC.
- Middleton, M. 2000. "Rock Singing." In *The Cambridge Companion to Singing*, edited by J. Porter, 28 - 41. Cambridge, UK: Cambridge University Press.
- Moylan, W. 2007. *The Art of Recording: Understanding and Crafting the Mix*. Oxford, United Kingdom: Focal Press.
- Paul, B. , and D. Huron. 2010. "An Association between Breaking Voice and Grief-related Lyrics in Country Music." *Empirical Musicology Review* 5 (2): 27 - 35.
- Poyatos, F. 1992. "The Audible-Visual Approach to Speech as Basic to Nonverbal Communication Research." In *Advances in Nonverbal Communication: Sociocultural, Clinical, Esthetic and Literary Perspectives*, edited by F. Poyatos, 41 - 58. Amsterdam: John Benjamins Publishing Company.
- . 1993. *Paralanguage: A linguistic and interdisciplinary approach to interactive speech and sound*. Amsterdam: John Benjamins.
- . 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.
- Rossing, T.D. 1990. *The Science of Sound*. Reading, Massachusetts: Addison-Wesley Publishing Company.
- Russell, J.A. 1980. "A circumplex model of affect." *Journal of Personality and Social Psychology* 39 (6): 1161–1178.

- Schimmack, U., and A. Grob. 2000. "Dimensional Models of Core Affect: A Quantitative Comparison by Means of Structural Equation Modeling." *European Journal of Personality* 14: 325-345.
- Smalley, Dennis. 1986. "Spectro-morphology and Structuring Processes." In *the Language of Electroacoustic Music*, edited by S. Emmerson. Hong Kong: The McMillian Press Ltd.
- . 1997. "Spectromorphology: explaining sound-shapes." *Organised Sound* 2 (2): 107 - 126.
- Tellegen, A., D. Watson, and L.A. Clark. 1999. "On the dimensional and hierarchical structure of affect." *Psychological Science* 10 (4): 297–303.
- 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*. Palgrave Mcmillian.
- Wescott, R.W. 1992. "Auditory Communication: Non-Verbal, Pre-Verbal, and Co-Verbal." In *Advances in Nonverbal Communication: Sociocultural, Clinical, Esthetic and Literary Perspectives*, edited by F. Poyatos, 25 - 40. Amsterdam: John Benjamins Publishing Company.
- Wilson, M. 2011. "Examining the effects of variation in emotional tone of voice on spoken word recognition." Master of Arts Masters Thesis, Psychology Cleveland State University. <https://engagedscholarship.csuohio.edu/etdarchive/569/>.
- Wishart, T. 1996. *On Sonic Art*. Amsterdam, The Netherlands: Harwood Academic Publishers.
- Zbikowski, Lawrence M. 2009. "Music, Language, and Multimodal Metaphor." In *Applications of Cognitive Linguistics [ACL]*, edited by Charles J. Forceville and Eduardo Urios-Aparisi. Berlin, New York: Mouton de Gruyter.
- <https://doi.org/10.1515/9783110215366.6.359>.