

# Emotional Tones and Emotional Texts: A New Approach to Analyzing the Voice in Popular Vocal Song

*Kristal Spreadborough*

NOTE: The examples for the (text-only) PDF version of this item are available online at:

<https://www.mtosmt.org/issues/mto.22.28.2/mto.22.28.2.spreadborough.php>

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

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 analyze due to the complexity and variety of the voice. This article 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 analyzing 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 and two new analytical parameters are introduced: the Affect Map and Cohesiveness. 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 voice in Kris Kristofferson’s 1970 song “Casey’s Last Ride.”

DOI: 10.30535/mto.28.2.7

*Received July 2021*

Volume 28, Number 2, June 2022  
Copyright © 2022 Society for Music Theory

---

## *Introduction*

[1] Humans have a rich palette 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. An artist’s spoken voice often plays a central role in their vocal aesthetic in popular vocal songs ([Lacasse 2010](#), 142). Because of this speaking-singing connection, the lived experience of speaking may heighten one’s sensitivity to vocal expression in popular 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 (Spreadborough 2018, chap. 4). One reason for this is the complexity of vocal sounds that occur in popular vocal songs. On a physiological level, this complexity is due to no two human bodies being exactly alike. Basic differences in the size and shape of biological structures impact one's overall vocal quality (Titze 1989). Lavan et al. explain, for example, that “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 (2019, 90). Furthermore, a single human vocal tract can be manipulated to produce a multitude of different vocal qualities (Lavan et al. 2019). Consider, for example, the differences in commonly heard vocalizations such as laughing, 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, 4). The social semiotic approach places our perception and lived experience of sound at the center of our understanding of sound. Others have sought to describe and analyze 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 analyzing acousmatic sounds that are not traditionally notated (e.g., Smalley 1986 and 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, chap. 2). Other approaches have sought to define and examine specific characteristics of sound (e.g., Lomax 1968), or have provided mechanisms to describe in detail the various sounds employed in vocal production (e.g., Wishart 1996, chap. 12). However, there is as yet no social semiotic framework for annotating and analyzing the sung voice.

[4] The goal of this current paper is to develop such a framework. In addition, this approach also allows the emotive content of tone quality to be compared with that of text (i.e., lyrical content), and to assess the implications of this relationship for a listener's overall emotional perception of the voice.

### The Analytical Approach

[5] **Example 1** shows a conceptualization of the analytical approach. The text in grey indicates the tools available for use at each level of the annotation and/or analysis. This conceptualization 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 conceptualization. In the case of popular vocal songs, this is even more apparent as idiosyncratic vocal sound qualities are often central to an 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 boundary enclosing “sound quality” and musical elements demonstrates that while these are definable attributes of a tone quality, they 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. It is worth mentioning here that it is due to this dialectic relationship between sound quality and musical

elements that tone quality more broadly is taken as the focus of the analytical technique presented in this paper.

[7] Text (i.e., lyrical content) is at the outermost layer in the conceptualization. Text includes both words and non-words that accompany the vocal performance. Emotionality of text is assessed against the **Affect Map** (discussed below). The tone/text relationship is the part of the conceptualization where the emotionality of tones and texts are compared and contrasted. This is achieved through the use of the tool **Cohesiveness** (discussed below).

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

[8] Before undertaking a discussion of the emotional models that underpin the approach presented in this article, 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 analyzing 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 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). 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 (Plazak 2011, 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 analyzing 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. 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 stylized form of speech, may sit anywhere along this spectrum. Serge Lacasse explores this connection in his 2010 analysis of Sia’s “Breathe Me.” Here, Lacasse approaches the sung voice as “a stylized 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 (143).

[12] This article will first define my two new parameters for analysis: the Affect Map and Cohesiveness. Next, my system for annotating and analyzing Tone Quality Features will be presented. Following this, an overview of the approach to textual analysis will be given. Finally, the analytical approach is applied to Kris Kristofferson’s 1970 song “Casey’s Last Ride” as a case study to demonstrate the potential for application.

### *Defining the Parameters of Analysis*

[13] This section outlines 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 is discussed first because it includes a discussion of the emotional models which underpin this paper. 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 and 1997).

## The Affect Map

[14] Before emotion *per se* is examined, it is important to consider the *locus* of emotion. In doing so, I draw 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 herein. According to Evans and Schubert, the locus of emotion refers to how listeners experience emotion in music—either by recognizing emotion expressed by the music, known as “perceived” or “external” locus of emotion; or through feeling a subjective response to the music, known as “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). Russell’s oft-cited dimensional model (1980) considers how emotion can be understood through two systems, valence (pleasure–displeasure) and arousal (activation–deactivation) (Eerola and Vuoskoski 2011). Another dimensional mode proposed by Tellegen, Watson, and Clark (1999) considers 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 Vuoskoski (2011). Eerola and Vuoskoski found that participants were able to rate emotional musical stimuli consistently and accurately 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 rate ambiguous emotions when using the dimensional model more accurately. 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 Vuoskoski 2011, 41)

[17] Unlike emotions, which tend to be short-lived affective states with words that 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). Hunter and Schellenberg (2010) also note that mood in music is a more diffuse experience that is typically not directed at a target (as is the case for emotion). In this article, I propose the use of an “Affect Map” 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) (Example 2).

[18] I use the term “affect” since it typically refers to an overall reaction 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, sec. 5.8). Since I wish to refer to both moods and emotions in my analysis, I adopt the term affect to describe the relationship between the two. No predefined taxonomy of emotions or moods is used in the Affect Map for two reasons. First, as summarized by Hunter and Schellenberg (2010), there is no single theory of what kinds of moods and emotions music evokes and conveys. 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 color 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 valence 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 to represent emotions of happiness and tenderness, thus creating a mood of contentment. Happiness in this example is positive (pleasant valence), of moderate energy (moderate arousal tension), and quite alert (high arousal energy). Tenderness in this example is moderately positive (moderately pleasant valence), relaxed (low arousal tension), and quite sedate (lower arousal energy). Taken together, these two emotions create a mood of contentment—moderate valence, 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 valence, arousal energy, and arousal tension that make up each mood and emotion.

### Cohesiveness

[21] Cohesiveness, shown in **Example 3**, is a tool developed for the analytical technique presented in this paper. It allows the analyst 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. 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 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 2010, 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 analyzing 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 of vocal and linguistic expression. For example, if one was to assess the vocal timbre and lyrics of Country Joe and the Fish’s song “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.

### *Annotating and Analyzing Tone*

[23] This section presents a classification system, called the Tone Quality Features, for annotating and analyzing 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 is also smooth and soft) (van Leeuwen 1999, 129). It also emphasizes the examination of sound not only in terms of what it “expresses” or “represents,” but also how it “affects us” (van Leeuwen 1999, 128). The tone quality features identified here are drawn from existing social semiotic approaches to voice quality (van Leeuwen 1999, Ngo and Spreadborough 2021). I have expanded on these features by including a system for describing onsets, something which is not addressed in existing social semiotic approaches to voice quality. I have also

provided an examination of the emotional implications for each Tone Quality Feature and developed a system for annotating 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 symbolized as  $\curvearrowright$ ).<sup>(1)</sup>

[24] Being able to succinctly and consistently describe acoustic features of a tone quality has two benefits. First, such descriptions help 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’s quality, then, also lends tone quality analysis a level of clarity which may not otherwise be available.

[25] Second, I also develop a system for discussing tone quality, which is a first step in analyzing how emotion is conveyed. For example, my previous research has suggested that, due to our lived experience of sounds, Tone Quality Features may be associated with particular emotions such as roughness and negative emotions (Spreadborough 2018, chap. 6). 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.

[26] A number of distinct features make up tone quality. Example 1 demonstrates two components, sound quality and musical elements, separated by a broken line. The broken line indicates that musical elements and sound quality are not discrete, but they are dialectically related (Fairclough 2001, 234). For example, some of the features within the Tone Quality Features constitute “musical elements.” One example of this is dynamic. On the other hand, a feature such as breath constitutes sound quality. However, dynamic and breath remain in dialogue because it is easier to produce more breathy sounds at softer dynamics than at louder ones. It is for this reason that the features described below are not categorized into discrete, separate lists for sound quality and musical elements—because each feature, while 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 dynamics, range, and vibrato). Thus, in the Tone Quality Features I describe 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 and duration are not included in this analytical approach.

## Level of Annotation and Analysis

[28] The level of annotation and analysis I apply to musical examples depends on the feature being analyzed. **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, 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 analyzed 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 analyzed is shown by rectangles.

## Tone Quality Features

[29] **Example 5** shows Tone Quality Features explored in this section as well as their corresponding graphical representation. Each feature can be annotated by using either the scale or by a discrete symbol. For example, a tone quality may be 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:  $\Delta\Delta\Delta\bullet$ ,  $\Delta\Delta$ ,  $\Delta$ ), or by utilizing one of three discrete symbols (e.g., very rough:  $\Delta\Delta\Delta$ , very smooth:  $\Delta$ , moderate:  $\Delta\Delta$ ). Onsets are not represented on a scale but are represented in terms of discrete symbols. This is because I consider the four onset types given in Example 5 sufficient for capturing the variation in the tone quality of an onset.

[30] The Tone Quality Features are not considered absolute, fixed points. Rather they are contextual—whether that be relative to 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 (Ngo and



Spreadborough 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 annotated on each of the features in Example 5), but also 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 (e.g., [Juslin and Laukka 2003](#)). 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 these basic emotions, I propose that irregularity could also be associated with nervousness and anticipation, which are not necessarily negative but are states of uncertainty. Therefore, high levels of variability would signify uncertainty, while low levels would signify certainty. In terms of the placement of variability on the Affect Map, extremely high levels of variability would be placed on the three dimensions as follows: more unpleasant (valence), either awake or tired (arousal energy), and more tense (arousal tension).

#### Onset

[32] Onset forms the initial part of the note and is important for the listener's ability to recognize timbre (see, for example, Saldanha and Coroso, as cited in [Erickson 1975](#), 61). I draw on three different classes of onset defined by Jo Estill ([McDonald Klimek, 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 **Example 6**.

#### 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 change and develop during the sustain. In developing the Tone Quality Features of sustain I draw primarily on theories of social semiotics of sound, specifically on the voice qualities outlined in van Leeuwen (1999, 129–141). Using my own findings ([Spreadborough and Anton-Mendez 2019](#), [Spreadborough 2018](#)), these features are expanded 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 sound. Rather, all Tone Quality Features are dialectically related—a sound is never just breathy, it is always a combination of all available Tone Quality Features. The analytical techniques presented here, especially the real time annotations, are designed to account for this by capturing different configurations of tone qualities simultaneously. The aim of this approach is to examine different configurations of tone qualities within the voice. These Tone Quality Features are described in **Example 7**.

#### Termination

[34] Termination, another tone quality feature, forms the end of the note. The duration of a termination ranges from very short to very long. **Example 8** describes these possible terminations and explores their emotional associations.

### *Analyzing Text*

[35] There are a range of approaches to text analysis in song. These range from explicit analysis of emotional words to more nuanced sentiment analysis utilizing theories such as Systemic Functional Linguistics (SFL) (e.g., [Ngo and Spreadborough 2021](#)). In this paper, I do not adopt a formal textual framework for analysis. I take this approach because text analysis is 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 one wishes. Since the purpose of this paper is to present a system for annotating and analyzing

tone, and a tool for comparing the emotionality of tone with the emotionality of 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 [Ngo and Spreadborough 2021](#)) and the analyses that result from such applications, provides ample opportunities for 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 emphasized in the musical phrase via their tone quality.

## *Application*

[36] 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 *Kristofferson*. **Example 9** shows the song timeline, characters, structure, and lyrics. The analysis of verse one will be presented first, followed by the analysis of chorus one.

[37] The equipoise of the song occurs at line one of verse one (see **Example 9**, 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 midpoint of the scale. Onsets are mostly simultaneous. While some Tone Quality Features tend towards the extremes of the scale (termination, breath, and vibrato), there is little variability in these features—that is, there is a consistency of tone quality which makes this a good example of tone quality equipoise.

### Verse 1

[38] The narrator introduces Casey in verse one. The lyrics progress from a sense of depression to a sense of unease. The lyrics in phrase one suggest a more relaxed, tired, and unpleasant emotive state (see **Example 10**) as they outline Casey's complicit descent into the subway. This is shown in **Example 10** where the Affect Map is used to represent the emotions in this verse. This configuration of arousal and valence is consistent with sadness. Phrase two moves to a more tense, awake, and moderately unpleasant emotionality (see **Example 10**) as the narrator describes Casey's isolation and the silent desperation of his descent. Such a configuration might suggest a moderate fear. The mixture of sadness and fear present in part one creates a sense of depression—portraying Casey's understanding of the desperateness, yet helplessness, of his situation.

[39] Phrase three conveys a lyrical message that is mostly tense, awake, and moderately unpleasant, which is typically associated with fear. 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. The movement of fear to more intense sadness creates a mood of unease in part two. The lyrics point to Casey knowing of the danger he is in, but suggest an inability to do anything to change the situation, as emphasized by the phrase "clicking of the turnstile" and "rattle of his chains"; the evocation of these quotidian sounds suggests that Casey is continuing in his established patterns.

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

[41] Onsets in part one are mostly aspirate and simultaneous (see **Example 11**). 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 lines 5–7. This increased use of glottal and creak onsets is shown in **Example 11**. 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 analyzed for their emotionality since 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 analyzed in **Example 11** since they are considered to be most neutral and also by-products of lyric pronunciation.

[42] A change in lyrical content accompanies this change in onset 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.

[43] 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 "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 11). 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.

[44] The glottal onsets in part two signify an emotional state that is tense, mostly awake, and moderately unpleasant (Example 12). This may be perceived as indicative of anger. The regular use 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 his situation may yet change. However, the second half of phrase 4 colors this message with a different emotion and mood.

[45] The second half of phrase 4 contains both a glottal and a creak onset (Example 11). The glottal onset on "clickin'" appears to perpetuate this sense of defiance. However, the placement of this glottal onset is different than in 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 creak onset on "rattle" may shed some light on this. This creak onset signifies a mostly relaxed, tired, and unpleasant emotive state. 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 12). 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). Such 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?

[46] 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 13). 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 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 14). This is not indicative of any particular emotive state, but it does create a mood of stability.

[47] The sustain in the second half of the verse, like the onsets, paints a different picture (Example 13). 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 realized 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 both suggest an emotive state that is more tense, awake, and unpleasant (Example 14). 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, the final line becomes gradually more lax, low, non-nasal and soft (Example 13), which contributes to the perception that Casey's emotional state has become mostly relaxed, tired, and unpleasant (Example 14). This is indicative of sadness and hopelessness (Example 14).

[48] 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 15). These weaker terminations occur at important structural points within the verse, on lines 4 and 8, which contribute to the two-part structure of verse one (Example 15). The tapering termination at the end of the verse is particularly interesting. 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. This reinforces the emotional message conveyed through tone quality: in verse one Casey has been through a journey of control, defiance, and finally hopelessness.

[49] The overall message delivered in the vocal line in part one of verse one is mostly aligned (**Example 16**). Tone quality features in this section do not tend towards the extreme of any emotion, but rather create a sense of steadiness. This moderate tone does not conflict with the text through which a general sense of depression is created. 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.

[50] The second half of the verse, on the other hand, presents a slightly different tone/text relationship (**Example 16**). The lyrics in the second half create a sense of unease. The tone quality, however, moves from a sense of defiance in lines 5–7 to a sense of hopelessness in line 8. While ending the verse with a tone quality which suggests hopelessness does create a sense of unease, in general the tone and text remain only somewhat aligned. 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?), and must also 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

[51] The narrator introduces The Woman in chorus one. **Example 9** shows the lyrics for this section. The lyrics switch to a first-person voice as The Woman pleads with Casey to stay with her. The lyrics of phrases 1, 3, and 4 suggest an emotional state that is mostly relaxed, moderately tired, and mostly unpleasant. This is shown in **Example 17** where the Affect Map is used to represent the emotions in this chorus. 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 throughout the 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. This configuration of arousal and valence is generally consistent with sadness.

[52] The lyrics of phrase 2, however, convey a different message. Here, the lyrics also suggest an emotional state that is relaxed and tired, conveying the emotional message implicitly. Yet, 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. This configuration of valence and arousal might be considered consistent with tenderness: The Woman still loves Casey even if he is making her sad. This mixture of sadness and tenderness creates an overall sense of despair (**Example 17**).

[53] The tone quality of this chorus in real time is annotated in **Audio-visual Example 2**. Aspirate and simultaneous onsets are used exclusively throughout this chorus except for phrase 4. Glottal and creak onsets appear in this phrase (**Example 18**). Similar to the analysis of onset above, aspirate and simultaneous onsets are not annotated since they tend to be byproducts of the pronunciation of lyrics. The use of such onsets does not suggest any highly charged emotional state. Rather, the tone is moderately tense, awake, and pleasant, creating a mood of stability (**Example 19**). Against this aural backdrop, the use of glottal and creak onsets in phrase 4 is quite salient.

[54] This change of onset is also accompanied by a change in lyrical message. In the first three phrases The Woman is speaking in declarative 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 19**). 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 19**). 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.

[55] The sustain and termination reflect this message too. The entire chorus is almost always plain, breathy, and non-nasal, but variation can be heard in other vocal features (see **Audio-visual Example 2**). Beginning in

phrase 2, variations in range, dynamic, and roughness are heard. Here, the word “just” is stronger and higher than the vocal quality in the previous line (**Example 20**). However, these variations are short lived, with a return to the low, soft vocal quality on the following words, “a kiss” (**Example 20**). But the delivery of “a kiss” is not as smooth as the preceding vocal quality has been. While this is not the first-time roughness has become present in this otherwise soft, low voice, it is the first time that roughness 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.”

[56] Phrases 3 and 4 also exhibit variations in sustain and termination. Phrase 3 begins with the personal pronoun “I” being delivered at a higher pitch than before. But as the phrase progresses, the sustain becomes lower and softer again (**Example 20**). The delivery of the final words “to please you” returns 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 20**). However, this is short lived as the final words “only stay a while” are delivered with a consistent roughness, persisting (for the first time) nearly the entire length of the phrase.

[57] Throughout the chorus, sustain and termination create an emotional message that is mostly relaxed, tired, and unpleasant (**Example 21**). This configuration is consistent with sadness (**Example 21**). 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.

[58] The overall message delivered in the vocal line in chorus one is mostly aligned (**Example 22**). Tone quality features suggest low arousal and negative valence, creating a mood of pain. On the whole this does not conflict with the text through which a general sense of despair is evident. Thus, the tone/text relationship in chorus one creates an affirming message for the listener. The emotionality of the vocal line is easy to access and understand.

[59] However, there are instances where vocal quality and lyrics are misaligned (**Example 22**). 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, prompting the listener to reassess the tone/text relationship in order to ascertain the overall message. The result is that at times the emotional message is not always obvious. Rather, this emotional message 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

[60] There is a noticeable contrast between the tone quality of verse one, which relates to Casey, and that of 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 his 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, while The Woman is the past. This dichotomy is established as early as the first few lines—Casey is following and seeing. This present is uncertain and dangerous, and this is reflected in the tone quality that is generally indicative of high arousal, unpleasant emotive states (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 immortalized—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.

[61] 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

[62] The goal of this paper is to present a new approach to analyzing the sung voice. This is achieved by presenting a new framework for analyzing 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. The analytical approach proposed in this paper is not intended to be prescriptive. Instead, my goal is to provide a consistent framework to annotate, analyze, and describe tone quality and its relationship with text. Others, in applying this 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 is an important avenue for research.

[63] The main implications for this approach relate to the annotation and analysis of the voice in popular vocal songs. 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, I present a framework on which future research can build in a variety of ways, including through musicology, multimodality, social semiotics, and psychology.

---

Kristal Spreadborough  
University of Melbourne  
Melbourne Connect (Building 290), Level 8  
700 Swanston Street  
Carlton, VIC, 3053  
Australia  
[kristal.spreadborough@unimelb.edu.au](mailto:kristal.spreadborough@unimelb.edu.au)

---

## Works Cited

- Campbell, Murray, and Clive Greated. 2001. "Loudness." In *Grove Music Online*.  
<https://doi.org/10.1093/gmo/9781561592630.article.17030>. Accessed December 15, 2020.
- Carter, Tim. 2001. "Word-Painting." *Oxford Music Online*.  
<https://doi.org/10.1093/gmo/9781561592630.article.30568>. Accessed August 24, 2018.
- Eerola, Tuomas, and Jonna 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>.
- . 2013. "A Review of Music and Emotion Studies: Approaches, Emotion Models, and Stimuli." *Music Perception* 30 (3): 307–40. <https://doi.org/10.1525/MP.2012.30.3.307>.
- Erickson, Robert. 1975. *Sound Structure in Music*. 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, Norman. 2001. "The Discourse of New Labour: Critical Discourse Analysis." In *Discourse as Data: A Guide for Analysis*, ed. Margaret Wetherell, Stephanie J.A. Taylor and Simeon J. Yates, 229–66. Sage Publishing.

- Frith, Simon. 1998. *Performing Rites: On the Value of Popular Music*. Harvard University Press.
- Haynes, Bruce, and Peter Cooke. 2001. "Pitch." In *Grove Music Online*.  
<https://doi.org/10.1093/gmo/9781561592630.article.40883>. Accessed December 15, 2020.
- Heidemann, Kate. 2016. "A System for Describing Vocal Timbre in Popular Song." *Music Theory Online* 22 (1).  
<https://doi.org/10.30535/mto.22.1.2>.
- Hunter, Patrick G., and E. Glenn Schellenberg. 2010. "Music and Emotion." In *Music Perception*, ed. Mari Riess Jones, Richard R. Fay, and Arthur N. Popper, 129–64. Springer Handbook of Auditory Research, vol. 36. Springer. [https://doi.org/10.1007/978-1-4419-6114-3\\_5](https://doi.org/10.1007/978-1-4419-6114-3_5).
- Huron, David. 2015. "The Other Semiotic Legacy of Charles Sanders Peirce: Ethology and Music-Related Emotion." In *Music, Analysis, Experience: New Perspectives in Musical Semiotics*, ed. Constantino Maeder and Mark Reybrouck, 185–208. Leuven University Press. <https://doi.org/10.2307/j.ctt180r0s2.17>.
- 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>.
- Lacasse, Serge. 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*, ed. Anne Danielsen, 141–58. Routledge.  
<https://doi.org/10.4324/9781315596983-11>.
- Lavan, Nadine, A. Mike Burton, Sophie K. Scott, and Carolyn McGettigan. 2019. "Flexible Voices: Identity Perception from Variable Vocal Signals." *Psychonomic Bulletin & Review* 26: 90–102.  
<https://doi.org/10.3758/s13423-018-1497-7>.
- Laver, John. 1980. *The Phonic Description of Voice Quality*. Cambridge University Press.
- Lomax, Alan. 1968. *Folk Song Style and Culture: A Staff Report on Cantometrics*. American Association for the Advancement of Science.
- McDonald Klimek, Mary, Kerrie B. Obert, Kimberly Steinhauer, and Think Voice International. 2005. *Estill Voice Training, Level Two: Example Combinations for Six Voice Qualities*. Estill Voice Training Systems International.
- Middleton, Richard. 2000. "Rock Singing." In *The Cambridge Companion to Singing*, ed. John Potter, 28–41. Cambridge University Press. <https://doi.org/10.1017/CCOL9780521622257.004>.
- Moylan, William. 2015. *The Art of Recording: Understanding and Crafting the Mix*. 2nd ed. Focal Press.
- Ngo, Thu, and Kristal Spreadborough. 2021. "Exploring a Systemic Functional Semiotics Approach to Understanding Emotional Expression in Singing Performance: Implications for Music Education." *Research Studies in Music Education* (December). <https://doi.org/10.1177/1321103X211034694>.
- Paul, Brandon, and David Huron. 2010. "An Association between Breaking Voice and Grief-Related Lyrics in Country Music." *Empirical Musicology Review* 5 (2): 27–35. <https://doi.org/10.18061/1811/46747>.
- Plazak, Joseph. 2011. "Instrumental irony and the perception of musical sarcasm". PhD Thesis. Ohio State University.
- Poyatos, Fernando. 1992. "The Audible-Visual Approach to Speech as Basic to Nonverbal Communication Research." In *Advances in Nonverbal Communication: Sociocultural, Clinical, Esthetic and Literary Perspectives*, ed. Fernando Poyatos, 41–58. John Benjamins Publishing Company. <https://doi.org/10.1075/z.60.08poy>.
- . 1993. *Paralanguage: A Linguistic and Interdisciplinary Approach to Interactive Speech and Sound*. Amsterdam Studies in the Theory and History of Linguistic Science. Series IV, Current Issues in Linguistic Theory. John Benjamins Publishing Company. <https://doi.org/10.1075/cilt.92>.

- 
- . 2002. *Nonverbal Communication Across Disciplines*. Volume II: Paralanguage, Kinesics, Silence, Personal and Environmental Interaction. John Benjamins Publishing Company.  
<https://doi.org/10.1075/z.ncad2>
- Rossing, Thomas D. 1990. *The Science of Sound*. Addison-Wesley Publishing Company.
- Russell, James A. 1980. "A Circumplex Model of Affect." *Journal of Personality and Social Psychology* 39 (6): 1161–78. <https://doi.org/10.1037/h0077714>.
- Schimmack, Ulrich, and Alexander Grob. 2000. "Dimensional Models of Core Affect: A Quantitative Comparison by Means of Structural Equation Modeling." *European Journal of Personality* 14 (4): 325–45.  
[https://doi.org/10.1002/1099-0984\(200007/08\)14:4<325::AID-PER380>3.0.CO;2-I](https://doi.org/10.1002/1099-0984(200007/08)14:4<325::AID-PER380>3.0.CO;2-I).
- Smalley, Denis. 1986. "Spectro-morphology and Structuring Processes." In *The Language of Electroacoustic Music*, ed. Simon Emmerson, 61–93. Palgrave Macmillan. [https://doi.org/10.1007/978-1-349-18492-7\\_5](https://doi.org/10.1007/978-1-349-18492-7_5).
- . 1997. "Spectromorphology: Explaining Sound-Shapes." *Organised Sound* 2 (2): 107–26.  
<https://doi.org/10.1017/S1355771897009059>.
- Spreadborough, Kristal L. 2018. "Voices Within Voices: Developing a New Analytical Approach to Vocal Timbre by Examining the Interplay of Emotionally Valenced Vocal Timbres and Emotionally Valenced Lyrics." PhD Thesis, The University of New England, Armidale, Australia.  
<https://doi.org/10.6084/m9.figshare.7636886.v1>.
- Spreadborough, Kristal L., and Ines Anton-Mendez. 2019. "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* 47 (3): 407–19. <https://doi.org/10.1177/0305735617753996>.
- Tellegen, Auke, David Watson, and Lee Anna Clark. 1999. "On the Dimensional and Hierarchical Structure of Affect." *Psychological Science* 10 (4): 297–303. <https://doi.org/10.1111/1467-9280.00157>.
- Titze, Ingo R. 1989. "Physiologic and Acoustic Differences between Male and Female Voices." *The Journal of the Acoustical Society of America* 85 (4): 1699–1707. <https://doi.org/10.1121/1.397959>.
- van Leeuwen, Theo. 1999. *Speech, Music, Sound*. Palgrave Mcmillian. <https://doi.org/10.1007/978-1-349-27700-1>.
- Wescott, Roger W. 1992. "Auditory Communication: Non-Verbal, Pre-Verbal, and Co-Verbal." In *Advances in Nonverbal Communication: Sociocultural, Clinical, Esthetic and Literary Perspectives*, ed. Fernando Poyatos, 25–40. John Benjamins Publishing Company. <https://doi.org/10.1075/z.60.07.wes>.
- Wilson, Maura L. 2011. "Examining the Effects of Variation in Emotional Tone of Voice on Spoken Word Recognition." MA Thesis, Cleveland State University.  
<https://engagedscholarship.csuohio.edu/etdarchive/569/>.
- Wishart, Trevor. 1996. *On Sonic Art. A new and rev. ed.* Edited by Simon Emmerson. Harwood Academic Publishers. <https://doi.org/10.4324/9781315077895>.
- 

## Footnotes

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.

[Return to text](#)

---



## Copyright Statement

Copyright © 2022 by the Society for Music Theory. All rights reserved.

[1] Copyrights for individual items published in *Music Theory Online* (*MTO*) are held by their authors. Items appearing in *MTO* may be saved and stored in electronic or paper form, and may be shared among individuals for purposes of scholarly research or discussion, but may *not* be republished in any form, electronic or print, without prior, written permission from the author(s), and advance notification of the editors of *MTO*.

[2] Any redistributed form of items published in *MTO* must include the following information in a form appropriate to the medium in which the items are to appear:

This item appeared in *Music Theory Online* in [VOLUME #, ISSUE #] on [DAY/MONTH/YEAR]. It was authored by [FULL NAME, EMAIL ADDRESS], with whose written permission it is reprinted here.

[3] Libraries may archive issues of *MTO* in electronic or paper form for public access so long as each issue is stored in its entirety, and no access fee is charged. Exceptions to these requirements must be approved in writing by the editors of *MTO*, who will act in accordance with the decisions of the Society for Music Theory.

This document and all portions thereof are protected by U.S. and international copyright laws. Material contained herein may be copied and/or distributed for research purposes only.

Prepared by Fred Hosken, Editorial Assistant

