TIMBRE IN MUSICAL AND VOCAL SOUNDS: THE LINK TO SHARED EMOTION PROCESSING MECHANISMS

A Dissertation

by

CASADY DIANE BOWMAN

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Takashi Yamauchi
Committee Members,	Jyotsna Vaid
	Jayson Beaster-Jones
	Thomas Ferris
Head of Department,	Douglass Woods

December 2015

Major Subject: Psychology

ABSTRACT

Music and speech are used to express emotion, yet it is unclear how these domains are related. This dissertation addresses three problems in the current literature. First, speech and music have largely been studied separately. Second, studies in these domains are primarily correlational. Third, most studies utilize dimensional emotions where motivational salience has not been considered. A three-part regression study investigated the first problem, and examined whether acoustic components explained emotion in instrumental (Experiment 1a), baby (Experiment 1b), and artificial mechanical sounds (Experiment 1c). Participants rated whether stimuli sounded happy, sad, angry, fearful and disgusting. Eight acoustic components were extracted from the sounds and a regression analysis revealed that the components explained participants' emotion ratings of instrumental and baby sounds well, but not artificial mechanical sounds. These results indicate that instrumental and baby sounds were perceived similarly compared to artificial mechanical sounds. To address the second and third problems, I examined the extent to which emotion processing for vocal and instrumental sounds crossed domains and whether similar mechanisms were used for emotion perception. In two sets of four-part experiments participants heard an angry or fearful sound four times, followed by a test sound from an anger-fear morphed continuum and judged whether the test sound was angry or fearful. Experiments 2a-2d examined adaptation of instrumental and voice sounds, where Experiments 3a-3d used vocal and musical sounds. Results from Experiments 2a, 2b, 3a and 3b were analogous such that aftereffects occurred for the perception of angry and not fearful sounds in different

domains. Experiments 2c, 2d, 3c, and 3d examined if adaptation occurred across modalities. Cross-modal aftereffects occurred in only one direction (voice to instrument and vocal sound to musical sound) and this effect occurred only for angry sounds. These results provide evidence that similar mechanisms are used for emotion perception in vocal and musical sounds, and that the nature of this relationship is more complex than a simple shared mechanism. Specifically, there is likely a unidirectional relationship where vocal sounds can encompass musical sounds but not vice-versa and where motivational aspects of sound (approach vs. avoidance) play a key role.

ACKNOWLEDGMENTS

I would like to extend my gratitude to my committee chair, Takashi Yamauchi, as well as my committee members, Jyostna Vaid, Jayson Beaster-Jones and Thomas Ferris for their invaluable input throughout the course of this research.

Thank you to my colleagues, especially Na Yung Yu and Genna Angello for their friendship and support during my time at Texas A&M University. Thank you also to the many upstanding research assistants that helped me create sound stimuli and collect data, without which this work would not be possible.

Finally, thank you to my family, specifically my mother and sister, for their encouragement and support. To my husband and daughters, thank you for your unending patience and love.

TABLE OF CONTENTS

Page

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

INTRODUCTION

Speech and music are two of the most effective means to express emotion through sound; they provide the basis for everyday social interactions (Juslin & Laukka, 2003). The domains of music and speech share numerous similarities and at the sound level and structural level (Fedorenko, Patel, Casasanto, Winawer, & Gibson, 2009) where rule based systems that contain rhythmic and melodic structures govern sequences of sounds (Patel, 2009). In conjunction, research in vocal acoustic (Bachorowski & Owren, 2008), infant-directed speech (Schachner & Hannon, 2011; Byrd, Bowman, & Yamauchi, 2012), and laughter (Bachorowski, Smoski, & Owren, 2001) suggest the idea of a shared emotion processing mechanism between music and speech. Is there something special about the perception of emotion in these two domains compared to other sounds? This question is the main motivation for my dissertation research.

1.1. Background

Emotions serve as a main component of communication in both the music and speech domains. In this chapter, I will introduce work regarding the role of emotion in speech and music as well as the role that acoustic components play in emotion perception. Because the focus of the following experiments involved participants from a Western culture, and stimuli consisted of Western instruments (e.g., the flute or saxophone as compared to a sitar or bagpipe), I will not delve into a detailed discussion on the cultural differences between speech and music. A short discussion,

however, is still necessary to understand some subtle differences in how music and speech sounds are perceived.

1.2. Emotion in music

Emotions represent reactions to an event of significance; they produce changes in an organism and function to communicate action and reaction in a social environment (Scherer, 1995; Darwin, 1872). Many expressive modalities are important to emotion communication such as body position, facial features, and vocalization (Scherer, 1995). Communication of emotion is crucial to social relationships and survival (Ekman, 1992) and two effective resources for emotional communication are speech and music (Thompson, Schellenberg, & Husain, 2004; Gabrielsson & Juslin, 1996).

Plato describes in *The Republic* that melodies in different musical modes (e.g., major, or minor mode) evoke different emotions (Patel, 2009). Since Darwin (1872), adaptive characteristics of music have been examined, such as emotion regulation and social communication (Scherer, 1995; Juslin & Sloboda, 2001). One use of music for emotion communication in everyday life is to regulate mood, such that listening to a slow piece of music creates a sense of calmness or well-being (Sloboda & O'Neill, 2001; Patel, 2009). An essential question addressed in music and emotion studies is how music evokes emotions (Eerola & Vuoskoski, 2013). Many studies have endeavored to identify emotions induced by music, as well as the acoustic components that contribute to emotion perception.

In one of the first theories concerning music-emotion relationships Meyer (1956) suggested that affective responses to music consist of experiences of tension and

relaxation, not actual emotions. This tension and relaxation occurs when listeners' expectations about what will happen in a piece of music is either violated or fulfilled (Hunter, Schellenberg, & Schimmack, 2010). Another model of emotion in music addresses how humans understand expressed or intended emotions (Figure 1, Balkwill & Thompson, 1999). This model indicates that there are universal cues (e.g., tempo, timbre and complexity) that influence a listener's emotional response to music. A listener uses salient cultural cues in music to arrive at an understanding of musically expressed emotions for familiar music (familiar tonal system) and perceptual cues when music is not familiar (unfamiliar tonal system).

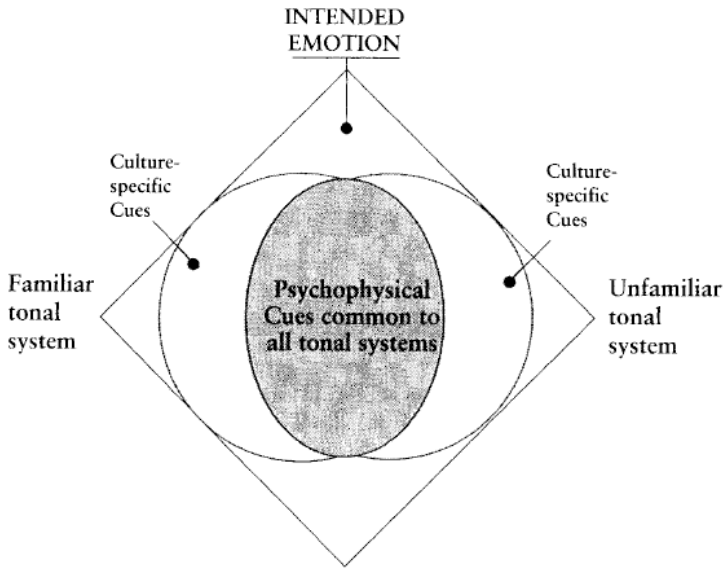


Figure 1. A model of musical emotion proposed by Balkwill and Thompson (1999). Each tonal system (familiar and unfamiliar) has its own distinct cultural cues that pertain to musically expressed emotions. Psychophysical cues that pertain to emotion are present within all tonal systems and provide an overlap of information that facilitates cross-cultural recognition of musically expressed emotion.

Models of emotion generally classify emotions in one of two ways, as basic or discrete. Basic or discrete emotions are commonly used in music as well as face and speech perception research (Bestelmeyer, Jones, DeBruine, Little, & Welling, 2010). Basic emotions are adaptive, and involve cognitive appraisal (Ekman, 1992); whereas, musical emotions are not adaptive or followed by direct external responses of a goal-oriented nature (Krumhansl, 1997). There is no current consensus on the best model to explain musical emotions, though behavioral, physiological, and neurological studies all indicate that listeners reliably have an affective response to music (Krumhansl, 1997; Gagnon & Peretz, 2003).

In summary, it is unclear whether music can convey specific emotions. Emotion studies in music have posited several theories ranging from expectation in music and chords (Hunter et al., 2010) to expressed and intended emotions (Balkwill & Thompson, 1999), to basic (Ekman, 1992) and dimensional emotions. These studies, however, have not demonstrated a firm consensus on the model of emotion that can best explain music.

1.3. Emotion in speech

Speech, like music, is a human universal. Speech works by use of a sensory-motor system, a conceptual-intentional system, and computational mechanisms which provide the capacity to generate an infinite number of expressions from a finite set (Hauser, Chomsky, & Fitch, 2002).

The transfer of information and the way speech is perceived depends on the meaning of the words spoken and the way something is said (e.g., prosody), which is often more revealing than what is actually said (Brück, Kreifelts & Wildgruber, 2012).

The information about a speaker's affective state is conveyed by the sound of the speaker's voice rather than vocabulary (Mehrabian & Ferris, 1967; Mehrabian & Wiener, 1967). For example, if a speaker is using a foreign language, humans are good at understanding the emotional state of the speaker simply by the tone and inflections of his or her voice (Pell, Monetta, Paulmann, & Kotz, 2009). Prosody is related to the typical way a person speaks and is mediated by modulations of parameters—pitch and timbre (Banse & Scherer, 1996; Kreifelts et al., 2013). For instance, when a speaker is happy, their voice rises in pitch and they increase volume and speak more quickly. In contrast, when sad, a speaker will use a quiet voice and a lower pitch at a slower pace (Banse & Scherer, 1996). Prosody is an important indicator of emotion in speech; however, other components of sound can provide information about speech and emotion, such as acoustic components of sound.

Perceptual experiments demonstrate that listeners are good at differentiating among emotion in speech (Banse & Scherer, 1996; Juslin & Laukka, 2003; see review in Juslin & Scherer, 2005). Voice-based cues, such as the tone of a person's voice when speaking or laughing, are powerful means to express emotion in spoken language (Kreifelts et al., 2013). In two studies, Bänziger, Patel, and Scherer (2014) showed that nonverbal vocal emotion communication is based on voice and speech features. Participants heard two sets of emotion utterances by German and French actors and were asked to rate the perceived voice and speech characteristics (loudness, pitch, intonation, sharpness, articulation, roughness, instability, and speech rate). Acoustic parameters were extracted from the voice samples and results showed that rater agreements were

high for most features (loudness, pitch, etc.). This indicates that the features used in the study were good descriptors of emotional speech and that this method can help identify other vocal features that are relevant for emotional communication (Bänziger, Patel, & Scherer, 2014).

There are several theories regarding emotion in speech. The source-filter theory of affect perception distinguishes how acoustic components provide information about emotional states (Kent, 1997; Bachorowski, 1999). Acoustic components commonly used in speech and emotion research are associated with the fundamental frequency of speech, which is perceived as vocal pitch (Bachorowski, 1999). Other important acoustic components in speech include jitter—which corresponds to variability in frequency –and shimmer, which corresponds to variability in amplitude. These components may be important for understanding emotional speech when taking into consideration other cues such as facial expression. For example, a sentence may sound different when a speaker is smiling in contrast to frowning (Bachorowski, 1999).

While music has been a pervasive facet in almost every culture, there is an ongoing debate of which capacities are utilized for music in the human brain and which might be shared with other cognitive domains (McDermott & Oxenham, 2008). Often, questions address how the voice is functionally and perceptually different from music; is there overlap in the brain regions that perceive music and language, and are the components used to perceive emotion within the two domains similar? More specifically, what is the link between speech, music and emotion?

1.4. The effects of culture on music and speech

Speech and music studies have primarily focused on a listener's sensitivity to music or speech in their own culture (Balkwill & Thompson, 1999). Musical behaviors including perception and judgment are universal and highly diverse in their structure, roles, and cultural interpretation (Trehub, Becker, & Morley, 2015).

Musical scales provide an example of a difference in emotion perception between cultures where many cultures use a system of scales as a foundation for building music. For instance, one difference is based on the amount of "tonal material" present in each octave of a scale (Dowling, 1978). In Western music there are 12 pitches per octave where 7 are typically chosen to build a musical scale. In contrast, Indian classical music uses "microtones" which are based on 7 pitches from 22 possible pitches in each octave that are separated by approximately ½ semitone (Patel, 2007). In addition, scales can differ in terms of interval patterns –the way the notes in a scale are spaced. For example, Western scales have a difference of one or two semitones in an interval, rather than equally spaced interval as found in some Javanese music with five intervals of equal size. These differences effect how emotions are perceived in different cultures' music.

While this is a simple example, there are many other ways in which cultures might differ with regard to the perception of music and related emotions. These dissertation studies are not aimed to focus on the cultural aspects of music and speech; nonetheless, the study of a cultures' effect on the relationship between music and speech is a promising endeavor that could shed light on how music and speech function as a unit and individually.

1.5. Acoustic components

There are many common components in music such as tempo—how fast or slow music is—and complexity—which generally involve the number of elements perceived in a piece of music; other acoustic components include timbre and loudness (Behrens & Green, 1993; Gabrielsson & Juslin, 1996). These components create structure and are further defined by Balkwill and Thompson (1999) as any property of sound that can be perceived independently of musical experience, knowledge, or enculturation. Such musical components are often regarded as "universal" and are presumed to extend beyond cultural contexts.

Acoustic components are the combined set of features used to perceive sound. In the speech domain, we recognize the identity of a spoken word across different speakers and we recognize a familiar voice across a range of utterances (Bergeson & Trehub, 2007). Similarly, in the music domain, we recognize melodies across changes in key (i.e., transpositions) or changes in musical instruments (i.e., timbre). Acoustic components act as the building blocks of sound and serve to create structure.

1.5.1. What are acoustic components

Acoustic components of affective sounds have been investigated since the 1970s (see Scherer & Oshinsky, 1977). There are eight known acoustic components related to timbre: attack time, attack slope, zero-cross, roll-off, brightness, Mel-frequency cepstral coefficients, roughness, and irregularity. These acoustic properties contribute to the perception of timbre in music and are likely to influence emotion independently of

melody and other musical cues (Hailstone, et al., 2009), making them ideal to study both music and speech.

1.5.2. Acoustic components of timbre

Attack time is the time in seconds it takes for a sound to travel from an amplitude of zero to the maximum amplitude in a sound signal. Attack time is known to contribute to the perception of emotion in music (Gabrielsson & Juslin, 1996; Juslin, 2000; Loughran, Walker, O'Neill & O'Farrell, 2004), which suggests that features of timbre are capable of determining the emotional content of music (Hailstone et al., 2009). The related feature *attack slope* is the attack phase of the amplitude envelope (shape) of a sound, and is interpreted as the average slope leading to the attack time.

Attack time and attack slope are computed using the linear equation, $y = mx + b$. This is part of a sound's amplitude envelope where m is the slope of the line and b is the point where the line crosses the vertical axis (t=0). For example, in Figure 2 the horizontal segments below the x-axis indicate the time it takes in seconds to reach the maximum peak of each frame for which the attack time is calculated. The arrows in Figure 2 indicate the slope of the attack.

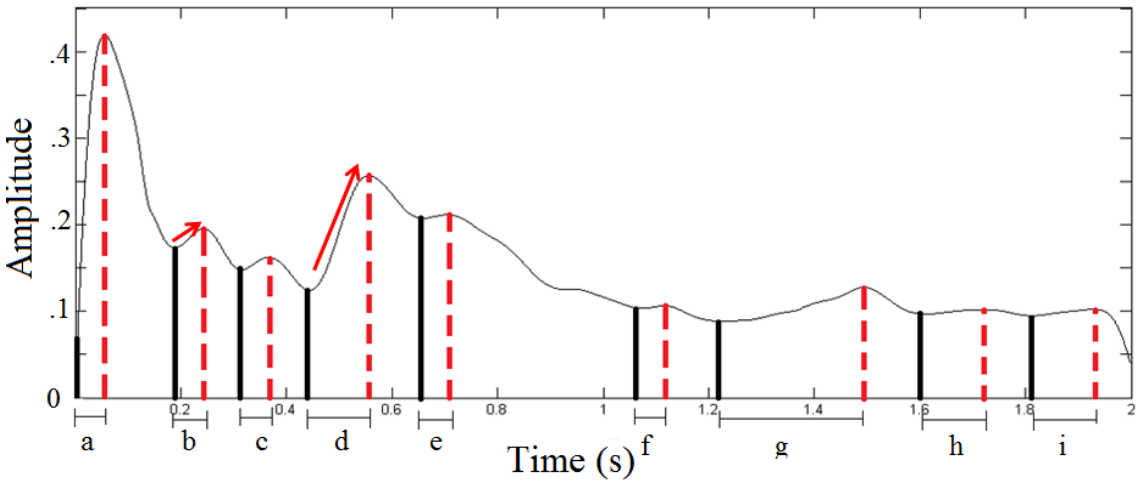


Figure 2. Attack time and attack slope of a waveform audio file. Sections *a* through *i* in the figure indicate separate attack times; this is the time in seconds from the vertical solid line, to the peak of the sound indicated by the vertical dashed line. The arrows indicates the duration (attack time) for which the attack slope is calculated.

Zero-cross is the number of times a sound signal crosses the x-axis for a frame (*t*) within a sound signal; this accounts for noisiness and is calculated using Equation 1 where *sign* is 1 for positive arguments and 0 for negative arguments. For frame *t*, *x*[*n*] is the time domain signal.

$$Z_t = \frac{1}{2}\sum_{n=1}^{N} | \, sign(x[n]) - sign(s[n-]) \tag{1}$$

Roll off is the amount of high frequencies in a sound signal. The roll-off frequency is defined as the frequency where the response is reduced by -3 dB. This is calculated using Equation 2, where *Mt* is the magnitude of the Fourier transform at frame *t* and frequency bin *n*. *Rt* is the cutoff frequency.

$$\sum_{n=1}^{R_t} M_t[n] = 0.85 * \sum_{n=1}^{R_t} M_t[n] \quad (2)$$

Brightness is the amount of energy above 1500 Hz and is related to spectral centroid. The term brightness is also used in discussions of sound timbres in a rough analogy to visual brightness. Timbre researchers consider brightness to be one of the strongest perceptual distinctions between sounds.

Roughness is a measure of sensory dissonance and is the perceived harshness of a sound; this is the opposite of consonance (harmony) within music or even single tone harmonics. Both consonance and dissonance are relevant to emotion perception (Koelsch, 2005). Roughness is calculated by computing the peaks within a sound's spectrum and measuring the distance between peaks. Dissonant sounds have irregularly placed spectral peaks as compared to consonant sounds with evenly spaced spectral peaks. Roughness is calculated using Equation 3, where *aj* and *ak* are the amplitudes of the components and *g* (*fcb*) is a 'standard curve.' This was first proposed by Plomp and Levelt (1965).

$$\rho = \frac{\sum_{j,k}^{n} a_j \cdot a_k \cdot g(f_{cb})}{\sum_{j}^{n} a_j^2} \quad (3)$$

Mel-frequency Cepstral Coefficients (mfccs) represent the power spectrum of a sound. This power spectrum is based on a linear transformation from actual frequency to the Mel-scale of frequency. The Mel-scale is based on a mapping between actual

frequency and perceived pitch as the human auditory system does not perceive pitch in a linear manner. Mel-frequency cepstral coefficients are dominant features used in speech recognition, voice-based affect detection, as well as some music modeling (Kwon, Chan, Hao & Lee, 2003; Logan, 2001; Neiberg, Elenius & Laskowski, 2006; Zeng, Pantic, Roisman & Huang, 2009). Frequencies in the Mel-scale are equally spaced and approximate the human auditory system more closely than linearly spaced frequency bands used in a normal cepstrum.

Irregularity is the degree of variation between peaks within a sound spectrum (Lartillot, Toiviainen, & Eerola, 2008). This is calculated using Equation 4, where irregularity is the sum of the square of the difference in amplitude between adjoining partials in a sound.

$$\frac{\sum_{k=1}^{N}(a_k - a_{k+1})^2}{\sum_{k=1}^{N} a_k^2} \tag{4}$$

All of these acoustic components work together to create the perception of timbre in a sound, which is essential for distinguishing two or more sounds with an identical pitch, duration and intensity. It is believed that brain mechanisms for processing timbre, and its acoustic components, are likely to have evolved for the representation and evaluation of vocal sounds (Juslin & Laukka, 2003).

1.5.3. Acoustic components in speech, music, and environmental sounds

Timbre is multidimensional (Caclin, McAdams, Smith, & Winsberg, 2005) and comprised of several acoustic components that help generate affect in a sound (Padova, Bianchini, Lupone, & Belardinelli, 2003). Temporal and spectral components (such as amplitude, phase, attack time, decay, spectral centroid, etc.) work simultaneously to influence the perception of timbre (Caclin, Giard, & McAdams, 2009; Caclin et al., 2005; Chartrand, Peretz, & Belin, 2008; & Moorer, 1977; Hailstone et al., 2009). These features are also essential for instrument recognition (e.g., Hajda, Kendall, Carterette & Harshberger, 1997). While the identity of a sound source may not be as important for a musical sound as it is for an environmental sound, its affective expression is of great significance (Scherer, 1995; Juslin & Laukka, 2003).

Eerola, Ferrer and Alluri (2012) showed that a dominant portion of valence and arousal could be predicted by a few acoustic components; such as, the ratio of high-frequency to low-frequency energy, attack slope and envelope centroid. Participants rated the perceived affect of 110 instrumental sounds that were equal in duration, pitch, and dynamics. Results showed that acoustic components related to timbre played a role in affect perception.

Scherer and Oshinsky (1977) used synthetic tone sequences of expressive speech with varied timbres and demonstrated that manipulating amplitude, pitch variation, contour, tempo, and envelope could explain variance in emotion ratings. Participants listened to one of three types of tone sequences created from sawtooth wave bursts and rated each sound on scales accounting for pleasantness-unpleasantness, activity-passivity

and potency-weakness and indicated if each sound was an expression of anger, fear, boredom, surprise, happiness, or disgust. While this showed strong effects of manipulating acoustic components of sound on emotion perception, this study did not address whether these components were related to timbre. Likewise, Juslin (1997) showed that listeners used similar acoustic components (e.g., tempo, attack time, sound level) to decode emotion in synthesized and live music performances. Results indicated that some acoustic components are related to specific emotions, but no direct comparison of components for timbre and emotion were made. Without this information, it is difficult to indicate how well timbre might explain emotion.

A study by Bowman and Yamauchi (in press) investigated the missing link between sound, timbre and emotion by examining whether particular acoustic components of sound that explain timbre also predicted particular categories of emotion (e.g., happy, sad, anger, fear or disgust; Ekman, 1992) in instrumental sounds. In two experiments, 180 synthetic sound stimuli were created from ten instruments (flute, clarinet, trumpet, tuba, piano, French horn, violin, guitar, saxophone and bell). In one experiment, participants received stimuli one at a time and rated the extent to which each stimulus sounded like its intended instrument (i.e., timbre judgment – how much a flute sounded like a flute). In another experiment, participants received the same sound stimuli and rated whether each of these stimuli sounded happy, sad, angry, fearful, and disgusting (i.e., emotion judgment). Analyses revealed that the acoustic components of regularity, envelope centroid, sub band 2, and sub band 9 explained ratings of timbre and emotion. The relationship between acoustic components and emotion judgments of basic

emotions was not uniform. For instance, for the instrumental sounds Sub band 7 (perceived activity in a sound) could predict anger, fear and disgust, but not sadness. Because shared acoustic components were found for timbre and emotion, it was speculated that timbre could be a more useful indicator for specific emotions (e.g., happiness or anger) rather than emotion in general.

Researchers have recently begun studying the relationship between emotion and timbre; yet several gaps in the literature exist. Effects of timbre are found in music and emotion studies, but the link between timbre and emotion is weak and there is lacking evidence for a conclusive set of acoustic components that explain both emotion and timbre (Coutinho & Dibben, 2012; Tuomos Eerola & Vuoskoski, 2013).

1.6. Emotion and timbre

Sounds are perceived and characterized by a number of attributes and components including pitch, loudness, duration, and timbre. Timbre is defined as the acoustic property that distinguishes two sounds of identical pitch, duration, and intensity; it is essential for the identification of auditory stimuli (Bregman, Liao & Levitan, 1990; Hailstone et al., 2009; McAdams & Cunible, 1992). When identifying a musical instrument, one uses timbre to tell the difference between a flute and guitar playing the same note. This quality of timbre allows a listener to identify individual instruments of an orchestra, and involves dynamic features of sound, especially onset characteristics (Grey & Moorer, 1977; Risset & Wessel, 1982).

1.6.1. What is timbre

Timbre is a feature of sound used to discriminate between two sounds that are identical in pitch and duration; it is often used when listening to a symphony to identify different instruments in the ensemble. The classic definition of timbre states that different timbres result from different amplitudes (of harmonic components) of a complex tone in a steady state (von Helmholtz, 1885), and /or the spectral distribution of energy of a sound. This definition illustrates the relationship between sound and timbre as it is a feature of sound, but does not adequately describe the acoustic components used create different timbres, and how these components overlap for the perception of emotion in sound.

Timbre is multidimensional and complex, and is made up of several acoustic components (Caclin et al., 2005). The complexity of timbre makes it difficult to study or measure on a single continuum such as low to high. Contrary to pitch, which relies on a tone‘s fundamental frequency and loudness, timbre relies on several parameters. A wide range of features from loudness and roughness (e.g., Leman, Vermeulen, De Voogdt, Moelants & Lesaffre, 2005) to mode and harmony (e.g., Gabrielsson & Lindstrom, 2010) can account for perceived emotions, but can these features explain the ability to perceive differences between sounds, such as the distinction between musical instruments or voices (i.e., timbre) (Patel, 2009)?

The main goal of most timbre studies has been to uncover the number and nature of its dimensions. A method most often used is multidimensional scaling (MDS) of dissimilarity ratings (Hajda et al., 1997; McAdams & Bigand, 1993). In studies using

MDS, listeners rate the dissimilarity between two stimuli, creating a dissimilarity matrix that undergoes multidimensional scaling to fit a perceptual timbre space. The dilemma with using this method is uncovering the acoustic components of timbre, and linking these to perceived emotions (McAdams, Winsberg, Donnadieu, De Soete & Krimphoff, 1995) in order to better understand how the two are related.

Overall, it is widely accepted that timbre is a quality of sound used to differentiate between two sounds that are equal in pitch, duration and intensity. For two reasons, however, this definition is flawed (Patil, Pressnitzer, Shamma & Elhilali, 2012). The definition of timbre is "negative." Instead of saying what timbre is, it is defined by what it is not. Second, the definition relies on a comparison between two sounds. The definition also does not encompass elements that are important to its meaning, such as the identification of out-of-sight predators, voices and speech of friends and family, or the recognition of musical instruments (Agus, Suied, Thorpe & Pressnitzer, 2012).

1.6.2. Timbre as a major component of emotion perception

Studies investigating the relationship between timbre and emotion have relied almost exclusively on the dimensional theory of emotion, which places emotions along continuous dimensions of *valence and activation* (Juslin, 2013). The problem with this is that everyday emotions are often perceived categorically (e.g., happiness, sadness, anger, surprise and fear; see Izard, 1977), guiding decisions for future behavior (Juslin, 2013). Evidence suggests that the ability to perceive different categories of emotion in music emerges early in cognitive development (Dalla Bella, Peretz, Rousseau, & Gosselin, 2001; Terwogt & Van Grinsven, 1991) and adults are able to decode emotions in music

categorically within just a few seconds of sounded notes (Peretz, Gagnon & Bouchard, 1998; Quinto, Thompson & Taylor, 2013). Results from over a hundred studies demonstrated that music listeners are generally consistent in their judgments of emotional expression (Juslin & Laukka, 2003). In addition, categorical emotions are easier to communicate than dimensional emotions in music (Gabrielsson & Juslin, 1996). While categorical emotions are recognized across cultures (Fritz et al., 2009), non-categorical emotions show low cross-cultural agreement (Juslin, 2013; Laukka, Eerola, Thingujam, Yamasaki, & Beller, 2013). The scope of this present research will make use of five basic emotions—happiness, sadness, anger, fear and disgust.

To summarize, acoustic features of sound can explain emotion (Eerola et al. 2012), yet it is not clear which model of emotion works best (dimensional versus categorical) to describe emotion. For instance, Schubert (2004) found acoustic features that could describe dimensional emotions (valence and arousal), but it is unknown how much his findings can be extended to specific emotions, such as sadness and fear, which are said to have similar valence but different levels of arousal. Furthermore, stimuli used in these studies were highly recognizable, for example, instrument sounds such as the flute or violin, which could have had a prior emotional association for listeners.

1.7. Problems with current music, speech and emotion studies

Despite the compelling findings, emotion processing underlying speech and music remains elusive due to three limitations. First, the majority of speech and music research has been conducted separately, not crossing domains. Only in the past several years have topics of interest in research expanded to include the perception of emotion in

music and speech (Juslin & Laukka, 2003; Patel, 2003). Second, the majority of the studies investigating emotional processing in these two domains is correlational, relying mainly on regression analysis (Byrd et al., 2011; Eerola et al., 2012; Juslin & Laukka, 2003). Regression analyses can determine what features of sound predict emotion ratings, but it only indicates an indirect associative relationship. Third, past literature does not make clear the effect of other facets of emotion such as discrete emotions or motivational aspects of emotion (e.g., approach versus avoidance). Due to these limitations, it is unknown whether the perception of emotion in speech and music is merely associative or structural, and a full understanding of emotion processing in speech and music is still unclear (Ilie & Thompson, 2006).

1.7.1. Research does not cross domains

Only recently have the domains of speech and music crossed paths. Many different expressive modalities are important to emotion communication such as body posture, facial features, and vocalization (Scherer, 1995); however, these domains remain largely separate. Because the domains of speech and music are similar with regard to several components, such as hierarchical structure, studying these domains together in terms of emotion perception is mutually beneficial.

People value music because of the emotions that it evokes. Musical abilities are important for the acquisition and processing of speech. To demonstrate, infants acquire information about words, word meaning, and phrases through the use of differing prosodic cues and acoustic components of sound (e.g., pitch and timbre). Across cultures, songs sung while playing with babies are fast, high in pitch and contain

exaggerated rhythmic accents, whereas lullabies are lower, slower and softer. Infants will use cues in both speech and music to learn the rules of a culture, which highlights the natural connection between speech and music. “Motherese” is a form of speech used by adults when interacting with infants and often consists of singing in a high-pitched, sing-song voice that mimics babies’ cooing to draw their attention and to help them learn (Fernald, 1989). Because infants begin life with the ability to make different sounds—first cooing and crying, then babbling —followed by word formation, full sentences and speech (Oller, 2000), motherese is a prime example of the use of music and sing-song qualities to aid in speech development. Music is crucial for both bonding with and soothing babies. Maternal speech has a number of features that can be considered musical and emotional, including higher pitch—which is associated with happiness—and a slower tempo, often associated with tenderness.

Like speech, the human capacity to create music is one of the most salient and unique markers that differentiates humans from other species (Miell, Macdonald, Hargreaves, & Cross, 2004). Byrd et al. (2012) showed that people’s ability to perceive emotion in infants’ vocalizations (e.g., cooing and babbling) was linked to the ability to perceive timbres of musical instruments. In one experiment, 180 pre-linguistic baby sounds were created by rearranging spectral frequencies of cooing, babbling, crying, and laughing made by 6 to 9-month-old infants. Participants listened to each sound one at a time and rated the emotional quality of the baby sounds. Results showed that five acoustic components of musical timbre (e.g., *roll off, Mel-frequency cepstral coefficient, attack time and attack slope*) could account for nearly 50% of the variation of the

emotion ratings made by participants. The results indicate that the same mental processes likely account for the perception of musical timbres and infants' prelinguistic vocalizations. While many similarities exist with regard to emotion perception, music and speech, most research in this area has largely been correlational, not demonstrating a causal relationship for the connection of emotion to music or speech.

1.7.2. Primarily correlational research

Vocal expression (i.e., the nonverbal aspects of speech, Juslin & Laukka 2003) and music (Gabrielsson & Juslin, 1996) are both nonverbal channels that rely on acoustic signals for communicating information. The suggestion of a close relationship between vocal expression and music has had a long history (von Helmholtz, 1863/1954, p. 371; Rousseau & von Herder, 1986); however, there is speculation about the relationship between these domains with no supportive empirical evidence.

Many studies have explored the link between the domains of music and speech, primarily using correlational analyses. Coutinho and Dibben (2012) examined how acoustic features of sound were related to emotion perception for speech and music. Listeners heard a 15 second music or speech sample and were asked to make an emotional rating based on a dimensional model of emotion (valence and arousal). Results showed that a set of seven psychoacoustic features: loudness, tempo/speech rate, melody/prosody contour, spectral centroid, spectral flux, sharpness, and roughness could explain both music and speech. These overlapping acoustic features for music and speech act to highlight the underlying similarities in neural processing. Again, these

results are only correlational and cannot distinguish whether there are shared mechanisms for emotion processing.

A review of 104 vocal expression and 41 music performance studies by Juslin and Laukka (2003) demonstrated the extensive nature of similarities between the two channels of communication. The focus of past studies has involved the accuracy with which discrete emotions were communicated to listeners and the way acoustic components were used to communicate emotion. The review explains that music is perceived as expressive of emotion, and is consistent with an evolutionary perspective of vocal expression of emotions (Juslin & Laukka, 2003). In summary, correlational studies are unsuitable to uncover the functional specificity underlying the music and speech domains (e.g., whether the same or different neural mechanisms mediate emotion processing in speech and music) (see Bestelmeyer et al., 2010 for exceptions, and Juslin & Laukka, 2003 and Eerola & Vuoskoski, 2013 for reviews).

1.7.3. Motivational salience

Though its effect on emotion perception of sounds is just beginning to be considered, motivational salience is not a new concept with regard to emotion. There is debate over what emotions are linked to approach and avoidance. Both approach motivation and avoidance motivation are governed by motives that orient or direct behavior toward or away from desired or undesired states (the *action-oriented view*; e.g., Carver, Sutton & Scheier, 2000; Eder, Elliot & Harmon-Jones, 2013). This is demonstrated in Wilkowski and Meier (2010) where faster approach movements were observed toward angry facial expressions showing that anger is related to approach

motivation rather than avoidance motivation. In contrast, Springer, Rosas, McGetrick and Bowers (2007) argued that angry faces were associated with heightened defensive activations (startle response/ avoidance). Other researchers also show that angry faces evoke approach or avoidance motivational reactions, depending on individual difference characteristics (Strauss et al., 2005). Regardless of the association of anger with approach or avoidance, this offers evidence that there are different sub regions of the amygdala that are sensitive to emotional cues from angry voices and indicates that more than one channel may be used to process emotion in vocal sounds.

1.8. Summary

While emotion research demonstrates the importance of emotional expression for communication, emotion research with regard to music and speech has not been studied jointly. Studies in speech and emotion have found that the communication of emotion does not depend solely on what is said, but how it is said (prosody), which is mediated by pitch and timbre (Banse & Scherer, 1996; Brück et al., 2012). It is yet unclear how these domains influence one another. Research on the perception of emotion in music suggests that music is used for mood regulation. Theories concerning musical emotions rely on the relationship between affect and experience. Meyer (1956) first proposed that affective responses to music were due to tension and relaxation, rather than actual emotions. In contrast Balkwill & Thompson (1999) found that psychophysical features—tempo, rhythm, complexity and pitch—are what listeners use to perceive emotion in music. Two current emotion theories that explain both music and speech are the discrete and dimensional approaches. Ekman (1992) proposed that basic emotions,

such as happiness, sadness, anger, fear, joy, disgust, sadness, shame and guilt are relevant in music and facial perception. The other currently held theory states that there are dimensional emotions, or emotions that vary along the continuous dimensions of valence and activation.

There are eight specific acoustic components of sound related to timbre that contribute to the perception of music and speech sounds. It is these acoustic components of sound that demonstrate an underlying relationship between emotional responses to music and speech. The acoustic components attack time, attack slope, zero-cross, roll off, brightness, Mel-frequency cepstral coefficients, roughness, and irregularity work together to create the perception of timbre in a sound. While Scherer and Oshinsky (1977) were some of the first to demonstrate that timbre has an effect on emotion ratings, Eerola et al. (2012) further demonstrated that timbre distinguishes valence and arousal in sound, and Juslin (1997) showed that listeners use acoustic components related to timbre to decode emotion in musical performances. Bowman and Yamauchi (in press) demonstrated that acoustic components of sound related to timbre explained timbre and emotion. Even with the research relating timbre and emotion, the link between these domains is weak; and there is lacking a definite set of acoustic features that explain both emotion and timbre (Coutinho & Dibben, 2012; Eerola & Vuoskoski, 2013).

CHAPTER II

REGRESSION STUDIES

2.1. Overview of experiments

In the following experiments the degree to which timbre-related acoustic components explained emotion perception of instrumental sounds, baby sounds and artificial mechanical sounds was examined. In Experiment 1a an audio synthesizer program was used to create 180 novel pseudo instrumental sounds by mixing frequencies from ten instrumental sounds (flute, clarinet, trumpet, tuba, piano, French horn, violin, guitar, saxophone and bell). Participants listened to and rated each sound for the affective qualities of happy, sad, anger, fear and disgust separately on a 1-7 Likert-type scale. In Experiment 1b, 180 pre-linguistic baby sounds were created by rearranging spectral frequencies of cooing, babbling, crying, and laughing made by 6 to 9-month-old infants. Participants listened to and rated each sound for the emotional qualities of happy, sad, anger, fear and disgust. In Experiment 1c (control condition), artificial mechanical sounds were used and were created in the same way as Experiments 1a and 1b. Participants rated the artificial sounds again for their emotional qualities. Experiment 1c acted as a control condition where the timbre related acoustic components were not expected to predict emotion ratings.

Eight acoustic properties of timbre: attack time, attack slope, zero-cross, roll off, brightness, Mel-frequency cepstral coefficients, roughness, and irregularity were extracted from all sound stimuli using MIRToolbox in Matlab (Lartillot et al., 2008). These acoustic properties are known to contribute to the perception of timbre in music

independent of melody and other musical cues (Hailstone et al., 2009). A random forest regression was applied to examine the extent to which these acoustic features could predict emotion ratings of instrumental, baby, and artificial mechanical sounds.

2.2. Experiments 1a-1c: instrumental, baby, and artificial mechanical sounds

2.2.1. Sound creation

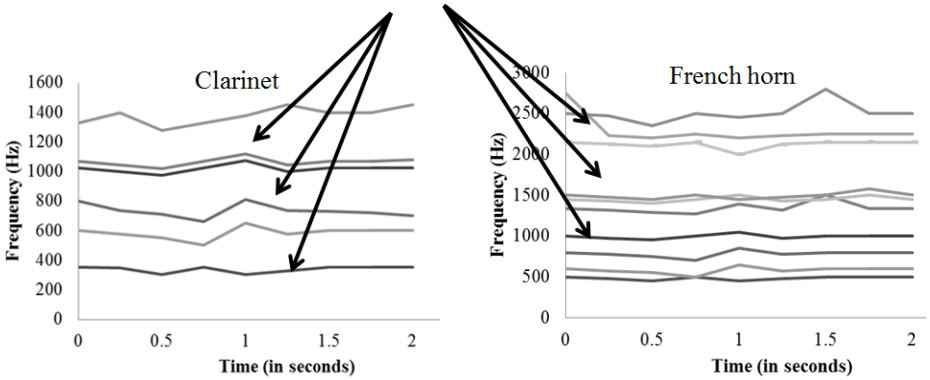
Novel instrumental (Experiment 1a), baby (Experiment 1b) and artificial mechanical sounds (Experiment 1c) were created for the experiments to increase the likelihood that there were no prior associations with emotion and the sound stimuli.

2.2.2. Creating instrumental sounds

"Pseudo" instrumental sounds were created (45 instrumental pairs X 4 emotions = 180 total sounds) from ten real instrumental sounds: flute, clarinet, alto saxophone, trumpet, French horn, tuba, guitar, violin, piano and bells (six professional musicians from the U.S. Army Reserve 395th band played the instruments at 440 Hz and a digital musical tuner was used for verification of pitch). Five undergraduate laboratory assistants were instructed to generate four different emotional sounds (happy, sad, angry and fearful) for each pair (45 pairs) of instrumental sounds using an audio editing and synthesis program SPEAR (Klingbeil, 2005). The synthesis program (SPEAR) applies fast Fourier transform analysis and decomposes each sound into amplitude and frequency components. Laboratory assistants created combination sounds from each pair of instrumental sounds by manually picking up frequencies from one sound (e.g., clarinet) and manually picking up frequencies from the other sound (e.g., French Horn), and mixing these frequencies to create a novel sound (Figures 3a and 3b). When creating

combinations, laboratory assistants were instructed to make sure that the combination sound still sounded like a mix between the two instruments in the given pair (e.g., the combination sound still sounded like a mix between the clarinet and the French horn).

3a. Step 1: Lab assistants select arbitrary frequencies from each sound in a pair



3b. Step 2: Randomly selected frequencies mixed to create a new "combined" sound

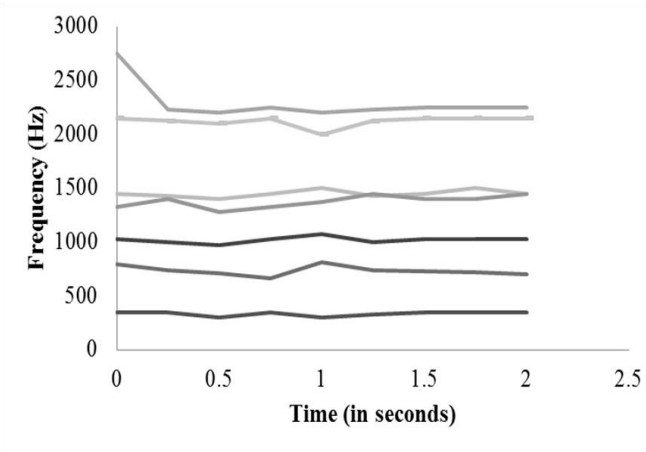


Figure 3. This figure illustrates the steps of stimuli creation. In step 1 frequencies were arbitrarily selected from each instrumental sound. In step 2, frequencies from two sounds were mixed. Lab assistants were instructed to maintain the sound identity of each instrument in the pair so that the new sound was an equal combination of the two instrumental sounds.

Laboratory assistants then modified the novel combined sound by manually shifting or deleting individual frequencies so that the sounds would convey happiness, anger, sadness or fear based on their own subjective judgments.

Prior to mixing, the sound amplitudes were normalized using the program Audacity (Version 1.3.4-beta) by utilizing the DC offset function where the mean amplitude of the sound sample was set to 0 to decrease any distortions or superfluous sounds not related to the stimuli. The instrumental sounds were then normalized by setting the peak amplitude to -1.0 dB

2.2.3. Creating baby sounds

The synthetic baby sounds were created in a similar manner as described for the instrumental sounds in Experiment 1a. Ten real infant sounds were used to create 180 synthetic baby sounds: five males and five females ranging from ages 6 to 9 months screaming, laughing, crying, cooing or babbling. Four sounds (one screaming boy, one crying boy, one screaming girl and one crying girl) were audio-recorded directly from two volunteer infants using an Olympic Digital Voice WS-400S recorder. The babbling and cooing sounds were taken from audio-files downloaded from a sound effects website (http://www.freesounds.org), and the laughing sounds were taken from files downloaded from YouTube (http://www.youtube.com).

These infant sounds were decomposed into spectral frequency components using SPEAR. Selected frequencies of one sound (e.g., a babbling sound of a boy) were mixed with selected frequencies of another sound (e.g., a cooing sound of a girl) and modified to convey one of four basic emotions—happy, sad, angry, and fearful. For each sound

pair (45 pairs in total) four sounds were created to sound like the emotion happy, sad, angry, or fearful, totaling 180 sounds. The sound stimuli were 2-5 seconds in length and normalized as in Experiment 1a, prior to mixing using the program Audacity (Version 1.3.4-beta).

2.2.4. Creating artificial mechanical sounds

Artificial mechanical sound stimuli were created in the same way as described in sections 2.1.1. and 2.1.2. for Experiments 1a and 1b. From 18 original recordings, 180 artificial sounds were created including bus exhaust, squeaking bicycle tires, and running AC units (see Table 1 for a list of sounds used to create combination sounds). None of the sounds included any speech or linguistic information. As in Experiments 1a and 1b, spectral frequency components and spectral frequencies of one sound (e.g., a bicycle tire) were mixed with spectral frequencies of another sound (e.g., bus exhaust) and modified to convey one of the four basic emotions—happy, sad, angry, and fearful. The sound stimuli were 2-5 seconds long and normalized prior to and after creation of each sound stimulus.

Table 1. Sounds used for stimuli in Experiment 1c.

Running air conditioning unit	Washing hands
Bicycle tires squeaking	Marker rolling on desk
Brakes squealing	Drawers opening
Bus exhaust	Clicking pen
Cart rolling in the library	Printer
Shades closing	Ripping paper
Compressor	Scratching on the wall
Crumpling paper	Shaking paper clips

2.3. Method

The procedure for each experiment was identical. Participants listened to sounds one at a time, and rated each sound on a 1-7 Likert-type scale for the emotions happy, sad, anger, fear and disgust. To obtain emotion ratings for individual sounds, emotion ratings were averaged over participants for each sound. Timbre related acoustic components were then extracted from each sound to examine the extent to which the components could account for emotion ratings given to individual sounds.

2.3.1. Participants

A total of 219 participants (73 male, mean age = 18.6, *SD* = 1.06; 146 female, mean age = 18.5, *SD* =.91) participated in Experiment 1a (instrumental sounds). Participants were randomly assigned to one of two groups that listened to 90 of 180 total sounds. A total of 145 participants (73 male, mean age = 18.6, *SD* = .99; 73 female, mean age = 18.7, *SD* = .94) participated in Experiment 1b (baby sounds). A total of 126 participants (56 male, mean age = 18.8, *SD* = 1.12; 70 female, mean age = 19.7, *SD* = .84) participated in Experiment 1c (artificial mechanical sounds). All participants took part in the experiments for course credit. Participants who were involved in one experiment (e.g., Experiment 1a) did not participate in the other experiments (e.g., Experiment 1b or 1c).

2.3.2. Materials

Stimuli for Experiments 1a, 1b, and 1c were 180 manually produced instrumental sounds, baby sounds, and artificial mechanical sounds, respectively.

2.3.3. Procedure

In Experiment 1a, 1b and 1c, participants were presented with sounds using customized Visual Basic software through JVC Flats stereo headphones. Each stimulus's maximum volume was adjusted and normalized. Participants listened to the stimuli, and rated each on five emotion categories, happy, sad, angry, fearful, and disgusting (Ekman, 1992; Johnson-Laird & Oatley, 1989). Each scale ranged from 1 to 7—1 being *strongly disagree* (the degree to which the stimuli, sounded like one of the five emotions), and 7 being *strongly agree*. Stimuli were presented in a random order. The rating procedure was the same for all experiments.

2.3.4. Design and analysis

Independent variables were predictors, or acoustic components (attack time, attack slope, zero-cross, roll off, brightness, Mel-frequency cepstral coefficients, roughness, and irregularity) extracted from the sound stimuli in each experiment. The dependent variables in Experiment 1a – 1c were the emotion rating scores averaged over participants for the 180 instrumental, baby, and artificial mechanical sounds, respectively.

To estimate the extent to which the acoustic components of timbre could predict emotion ratings, random forest (Liaw & Wiener, 2002) was applied. Random forest is a non-parametric method. It employs "ensemble" learning; 500 or more decision trees are formed by randomly selecting observations and variables. By aggregating "votes" cast by these random decision trees, the algorithm generates estimated likelihoods of a dependent variable. The prediction performance of the acoustic components was

measured by Out of Bag (OOB) cases—cases that were not used for training. Thus, our OOB prediction performance measure was equivalent to a boot-strap cross validation method (Breiman, 2001). To avoid overestimation of prediction performance, no parameter tuning was employed and default parameters implemented in the random forest R package (Liaw & Weiner, 2002) were applied in the analyses. To compare prediction performance, R^2 (i.e., 1-(SSE/SST)) was reported, which indicates the variance explained by the model.

2.4. Results

This section begins with an overview of the behavioral data from Experiments 1a (instrumental sounds), 1b (baby sounds) and 1c (artificial mechanical sounds) followed by results indicating how well acoustic features could explain emotion ratings in the instrument sound rating task (Experiment 1a), the baby sound rating task (Experiment 1b) and the artificial mechanical sound rating task (Experiment 1c).

2.4.1. Descriptive statistics

Figure 4 shows overall observations for each emotion for all sounds in Experiment 1a-1c. The boxplot in each figure represents the distribution of the 180 rated sound stimuli for each emotion. The whiskers of the boxplots indicate the variation of each rated emotion for the 180 sound stimuli and the median represents which emotions were rated the lowest or highest. In Figure 4a, the whiskers show that the ratings of the 180 instrumental stimuli are varied and range between 2.8 and 4.0, based on the median. Figure 4b demonstrates similar results for baby sound stimuli where there was similar variation in the data and the median ranges between approximately 2.5 and 4.75, with

more sounds rated as angry and least like the emotion happy. Figure 4c represents behavioral data for the artificial mechanical sounds where there was considerably less variation compared to instrumental or baby sounds. Sounds were rated as high in fear and anger and least like the emotion happy, where the median ranged between approximately 2.5 and 4. Overall there was good variation for emotion ratings of the sounds for both instrumental and baby sounds. The artificial mechanical sounds, however, were less varied in the ratings of emotion for the 180 sounds.

a.

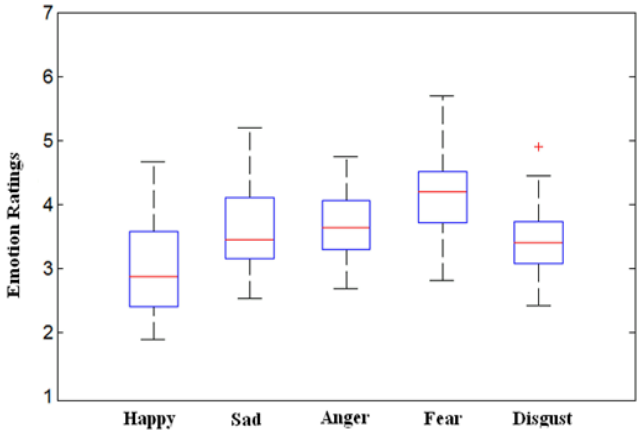
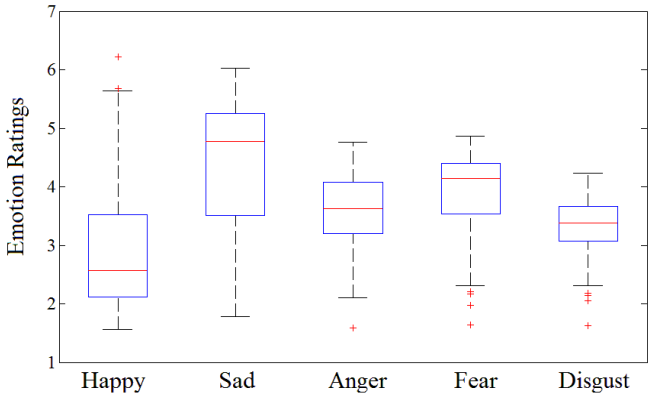


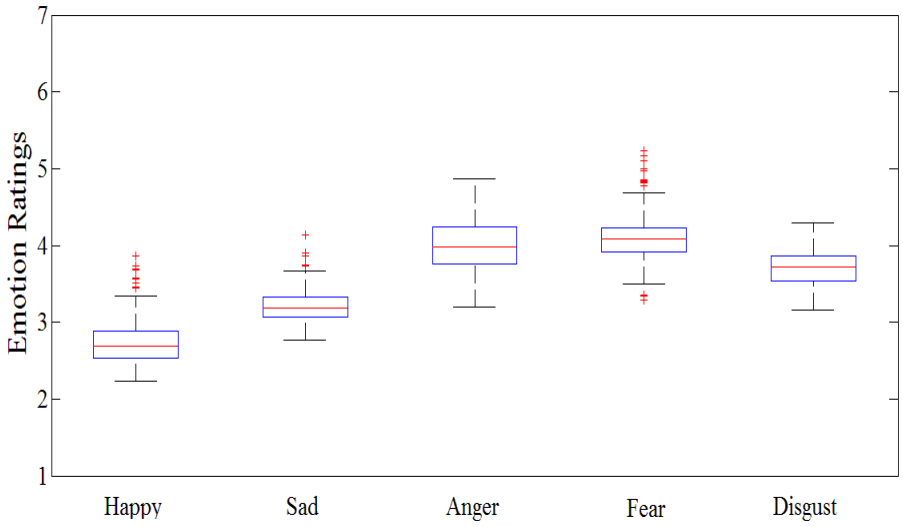
Figure 4. Boxplots of emotion ratings for (a) instrumental, (b) baby, and (c) artificial mechanical sounds. The center line of each box is the median, the edges indicate the 25th and 75th percentiles, and whiskers indicate extreme data points. Outliers are plotted outside of the whiskers.

Figure 4 continued.

b.



c.



2.4.2. Random forest regression analysis

Overall, the eight predictors could explain the instrumental and baby sounds well; however, the artificial mechanical sounds were not explained by as many of the

acoustic components. These results indicate a stronger link between music and speech sounds, compared to artificial mechanical sounds.

To assess how well the eight predictors (acoustic components) explained averaged emotion ratings of the instrumental sounds, percent variance, or R^2, was used; see the first row in Tables 2-4. Percent variance explains how much of the variance in emotion ratings was accounted for by the acoustic components used as predictors. In addition, importance scores of each predictor were assigned to the acoustic components. These scores were generated by the random forest algorithm and indicate the degree of contribution of individual features in the model.

For Experiment 1a (instrumental sounds), the results of the regression indicated that 42% of the variance in the emotion happy was explained by the eight acoustic features and 40% of the variance explained the emotion sad. The acoustic components accounted for 34% of the variance in the emotion anger and for the emotion fear the components explained 31% of the variance. Only 19% of the variance for disgust was explained by the predictors. The eight acoustic components related to timbre best explained the emotions happy, sad and anger for instrumental sounds. Overall, the predictors worked well to explain emotion ratings of the instrumental sound stimuli where the emotions happy and sad were explained better than other emotions. These results indicate that musical timbre is a good descriptor for emotion in instrumental sounds. Table 2 summarizes percent variance explained by the eight predictors for each emotion and shows importance scores for each of the eight acoustic components.

Table 2. Importance scores for instrumental sounds (Experiment 1a).

Percent Variance	42.13	40.00	33.50	31.15	19.10
	happy	sad	Anger	fear	disgust
attack time	4.15	4.31	2.64	3.88	2.06
attack slope	**12.74**	6.26	**4.49**	**6.75**	3.13
zero crossing	**11.02**	**11.48**	**4.28**	5.88	**4.06**
roll off	6.63	**11.77**	4.16	4.54	**4.25**
brightness	6.19	**8.61**	3.31	4.60	3.49
irregularity	**8.50**	5.36	3.61	**6.76**	3.05
mfcc	7.13	7.09	**4.87**	**6.92**	**3.92**
roughness	**25.54**	**9.10**	**8.36**	**15.91**	**6.28**

The first row is percent variance accounted for by the predictors for each emotion. The values in the table represent importance scores, or weighted values of the predictors

The results of the regression indicated that for Experiment 1b (baby sounds), the eight acoustic features explained over half, or 55%, of the variation in sad emotion ratings, see Table 3. Fear was the next best explained emotion by the predictors at nearly half, or 47.5% variance. Forty-five percent of variance in the emotion ratings for the emotion happy was explained by the eight predictors with 41.5% for anger and only 31% for the emotion disgust. The eight acoustic components related to timbre best explained the emotions sad, fear and happy for baby sounds. These results showed that, similar to instrumental sounds, the acoustic components worked well to explain emotion in baby sounds.

Table 3. Importance scores for baby sounds (Experiment 1b).

Percent Variance	45.33	55.37	41.53	47.50	31.37
	happy	sad	anger	fear	disgust
attack time	**20.67**	**21.80**	5.47	**9.95**	2.81
attack slope	14.13	13.27	5.55	6.57	3.57
zero crossing	**22.05**	**23.32**	**11.32**	**10.68**	**5.63**
roll off	**32.41**	**38.94**	**11.44**	**11.68**	**4.83**
brightness	**22.29**	**23.91**	**9.71**	**9.61**	**5.28**
irregularity	16.80	18.58	6.09	6.34	2.81
mfcc	14.05	16.18	**6.15**	6.30	2.52
roughness	16.53	13.86	5.20	8.50	**3.65**

[b.]The first row is percent variance accounted for by the predictors for each emotion. The values in the table represent importance scores, or weighted values of the predictors.

The results of the regression for Experiment 1c (artificial mechanical sounds) indicated that 35% and 34% of the variance in the emotions fear and happy were explained by the eight acoustic features, see Table 4. To a lesser degree anger and sad were explained by 29% and 22% variance, where disgust was not explained by the acoustic components. The results of the regression indicated that artificial sounds were not explained well by the eight acoustic components compared to either instrumental or baby sounds (see Figure 5). This result alone suggests that timbre could be a driving force for emotion processing for music and speech, but not for artificial sounds.

Table 4. Importance scores for artificial mechanical sounds (Experiment 1c).

Percent Variance	33.58	21.49	29.43	35.01	0
	happy	sad	Anger	fear	disgust
attack time	1.53	0.55	1.50	1.34	0.64
attack slope	**4.64**	**1.44**	2.13	1.68	0.92
zero crossing	**1.89**	0.90	**2.38**	**1.72**	**1.03**
roll off	1.64	**1.33**	**2.28**	**2.49**	**0.93**
brightness	1.48	**0.99**	2.26	**2.35**	0.86
irregularity	**2.04**	**1.43**	**2.63**	**4.94**	0.91
Mfcc	1.45	0.81	2.04	1.71	**0.92**
roughness	**1.80**	0.82	**4.38**	1.62	**1.05**

[c.] The first row is percent variance accounted for by the predictors for each emotion. The values in the table represent importance scores, or weighted values of the predictors.

Generally, predictors that explained both instrumental and baby sounds, did so at a much higher percentage (R^2) compared to artificial sounds. Moreover, the predictors that worked well to explain instrumental and baby sounds had much higher importance scores, where those predictors that could also explain mechanical artificial sounds had much lower importance scores. This discrepancy in the weights of importance scores also shows that the predictors did not work as well to explain emotion in the artificial sounds compared to the instrumental and baby sounds. The predictor that worked well to explain both instrumental and baby sounds was zero crossing. Because it worked well to explain both types of sounds, this particular acoustic component could be more predictive of emotion in general in other types of sounds. See Figure 5 for a comparison of R^2 values for the instrumental, baby, and artificial mechanical from the random forest regression, broken down by emotion.

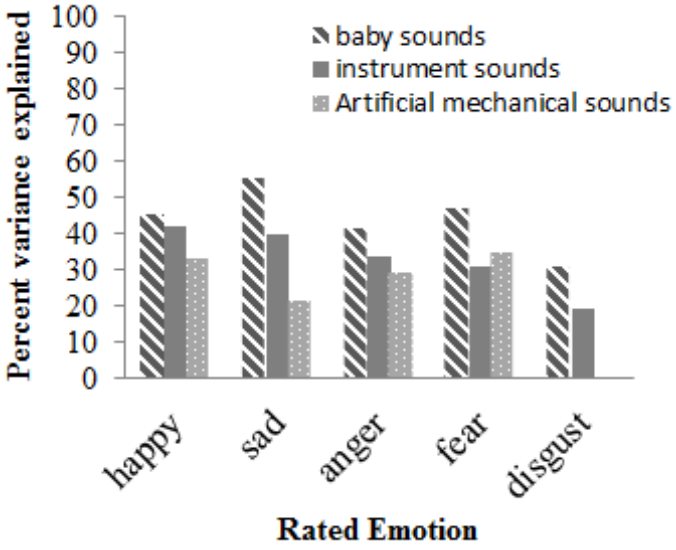


Figure 5. R^2 values for each emotion for instrumental (striped bars) baby (solid bars) and artificial mechanical (dotted bars) sounds.

2.5. Discussion

Experiments 1a-1c examined whether acoustic predictors of timbre could explain emotion ratings in instrumental, baby and artificial mechanical sounds. The goal was to identify timbre-related acoustic components that could explain emotion perception in baby, instrumental, and artificial mechanical sounds. Overall, results from Experiments 1a-1c demonstrated that the acoustic components worked much better to explain emotion ratings from instrumental and baby sounds compared to artificial mechanical sounds. Because sounds such as squeaking bicycle tires and car exhaust were not explained well by the timbre components, this indicates that those sounds related to music (instrumental sounds) and speech (baby sounds) are special in comparison to other sounds.

Music, speech, and even ambient sounds carry emotional information that is transmitted via the acoustics of the sound and then decoded by the audience of a concert,

another person, or an artificial intelligence system (Weninger, Eyben, Schuller, Mortillaro, & Scherer, 2013). Recent work in affective computing has demonstrated similarities for music, speech and other types of sounds (Drossos, Floros & Kanellopoulos, 2012; Isabelle Peretz, Radeau, & Arguin, 2004; Roesch et al., 2011); however, there is not yet a computational model that can account for general affect perception in sound. Results from this study demonstrated the interconnectedness between instrumental and baby sounds with regard to emotion and acoustic components. Because vocal sounds carry affective and semantic information, and acoustic features used for emotion perception overlapped with that of instrumental sounds, perhaps these sounds communicate emotions using a shared mechanism. Generally, if music and speech did co-evolve and instruments were made for emotion communication (perhaps by mimicking speech sounds), then instrumental sounds may act as a go-between on a continuum of emotional salience which ranges from mechanical sounds to speech.

Though results indicated a relationship between emotion perception of instrumental and baby sounds, some limitations exist. For example, acoustic components may not have explained the artificial mechanical sounds to a great degree due to a small variance in the emotion ratings of the mechanical sounds. The boxplot for rated emotion of the 180 artificial mechanical sounds indicated a very small range for emotion ratings of these sounds, which could limit how well the acoustic components worked to explain these sounds.

Overall, baby sounds were explained better than instrumental sounds by the acoustic components. It is plausible that these sounds are perceived as an intermediary

between speech and mechanical sounds. For example, speech sounds are produced by passing air over the vocal chords, however, instrumental sounds are produced by a person acting on an object (e.g., the flute) to create a sound and convey emotion. Mechanical sounds, however, are not produced by humans acting on an object in order to convey emotion (e.g., a pencil rolling on a desk does not convey anger). Thus, in the perception of emotion of different types of sounds (e.g., baby versus mechanical) there potentially exists a gradation of emotion perception that is determined by how a sound is produced.

CHAPTER III

ADAPTATION STUDIES

3.1. Why study adaptation

Although recent research reveals a link between timbre, emotion, and the music and speech domains, it predominately relies on correlation and regression analysis (Byrd et al., 2011; Eerola et al., 2012; Juslin & Laukka, 2003). What is lacking is empirical research to show that there is a causal link between musical and vocal sounds.

The perception and recognition of signals conveying affect (e.g., from faces or voices) is important and used for everyday social functioning (Bestelmeyer et al., 2010). In the auditory domain, nonverbal signals are crucial in communicating emotional information (Wallbott & Scherer, 1986). Previous research demonstrated perceptual aftereffects for both emotionally expressive faces and vocal sounds; however, the extent to which these aftereffects can cross modalities—voice to instrument—has not been studied. By investigating adaptation in the domains of speech and music we can assess the extent to which mechanisms for emotion processing in the two domains overlap.

Adaptation is a process during which continued exposure to a stimulus results in a biased perception toward opposite features of the adapting stimulus (Bestelmeyer et al., 2010; Grill-Spector et al., 1999). MacLin, Nelson and Webster (1996) showed that extended exposure to distorted faces caused non-manipulated faces to appear distorted in the opposite direction of the adapting stimulus. Often, adaptation paradigms are utilized to probe functional specificity of neural populations (Bestelmeyer, Maurage, Rouger, Latinus & Belin, 2014).

A classic example of adaptation is the color aftereffect, where an observer perceives a green square after-image following adaptation to a red square (Clifford & Rhodes, 2005). While color aftereffects are due to the adaptation of color-opponent cells in the retina, experiments have also shown adaptation aftereffects for high-level visual stimuli such as faces, across dimensions such as identity, gender, race and expression (Fox & Barton, 2007; Leopold, O'Toole, Vetter & Blanz, 2001; Webster, Kaping, Mizokami & Duhamel, 2004). For example, Bestelmeyer et al. (2010) demonstrated that auditory adaptation to angry vocalizations causes voices at test to be perceived as more fearful, and vice versa.

Adaptation research shows that neurons respond to specific stimulus attributes and are active at early stages of information processing, particularly for high-level properties such as facial identity (Bestelmeyer et al., 2010; Grill-Spector et al., 1999; Leopold et al., 2001). Researchers interpret these aftereffects to mean that a recalibration of neural processes takes place in response to continuously updated stimulation (Bestelmeyer et al., 2010; MacLin et al., 1996), such that neurons are "worn out" from responding to an angry stimulus adaptor and then recalibrate so that an ambiguous sound at test is perceived as less angry.

Commonly, face adaptation studies use paradigms that involve morphed faces. Participants are shown a particular face during a short adaptation period, and then shown ambiguous test images created by morphing between two faces. Adaptation causes these subjects to respond such that the morphed images are less similar to the face they had viewed during the adaptation phase. This aftereffect is attributed to a reduction in neural

responses evoked by the adapting face (Huber & O'Reilly, 2003). Following the adaptation phase, responses in competing unadapted representations of faces are stronger than the response in the adapted representation (Leopold et al., 2001). These results suggest that adaptation methods are a useful and important means of uncovering the nature of the neural representations of faces and facial representations in the human visual system (Butler, Oruc, Fox & Barton, 2009; Rhodes, Brennan & Carey, 1987).

Webster and MacLin (1999) were the first to show that extended exposure to faces can also generate aftereffects. Adaptation to consistently distorted faces (e.g. expanded features) caused subsequently viewed unmanipulated faces to appear distorted in the opposite direction of the adapting stimulus (e.g. compressed features). This effect transferred to faces of different identities. In a study by Bestelmeyer et al. (2010) the visual perception of complex stimuli and faces show that nonlinguistic information in voices elicits auditory aftereffects. For example, adaptation to male voices causes a voice to be perceived as more female (and vice versa), and these auditory aftereffects are measurable even minutes after adaptation. This adaptation effect did not cross modalities. Adaptation effects were absent, both when male or female first names were used as stimuli and when silently articulating male or female faces were used as adaptors (Schweinberger et al., 2008).

Prolonged exposure to stimuli can also result in the opposite effect—sensitization. Sensitization results when an observer is repeatedly exposed, for instance, to an angry face and rates a subsequent face as angrier (Kandel & Siegelbaum, 2012, p. 1465). The exact interpretation of what causes sensitization is still unclear. Recent

behavioral and fMRI research points to the idea that sensitization is mediated by similar processes as adaptation and that sensitization may occur when stimuli serve a salient adaptive purpose (Frühholz & Grandjean, 2013). Frühholz and Grandjean (2013) demonstrated that angry vocalizations evoked changes in the brain such as an increased alertness, which caused sensitivity to emotional information that is important for adaptive behavior. Participants listened to four speech-like, non-word stimuli and rated prosody discrimination of voices (e.g., if the voice was neutral or angry) while recorded on fMRI. Results show sensitization where the bilateral superficial (SF) complex and the right laterobasal (LB) complex of the amygdala were sensitive to emotional cues from speech prosody that were similar to a melody in music. This offers evidence that anger, which has negative valence but approach motivation, is processed separately from fear, which has negative valence and avoidance motivation.

3.2. Instrument and voice

3.2.1. Overview of experiments: 2a – voice → voice, 2b – instrument → instrument, 2c – voice → instrument and 2d – instrument → voice

While the adaptation paradigm has been used to explore neural mechanisms underlying face perception, it is not yet clear if these aftereffects exist for processing other types of nonlinguistic auditory information, such as vocal and instrumental sounds. To empirically investigate the relationship between the speech and music domains, I focused on the link between voice and instrumental sounds. Voice and instrumental sounds were used as an initial starting point for studying speech and music because they are simple and lack some of the complex variables such as rhythm or prosody. By using

an adaptation paradigm designed by Bestelmeyer et al. (2010; 2014), I investigated the structural relationships between voice sounds and instrumental sounds and emotion.

In Experiment 2a, participants heard either an angry or fearful vocalization from the Montreal Affective Voices (Kawahara & Matsui, 2003) four times to elicit adaptation. Following this exposure phase, participants heard a test sound from a morphed continuum of the same voice sounds from the MAV (adapted to voice→tested on voice). Experiment 2b was similar to Experiment 2a, except participants heard instrumental sounds at exposure and test phases (adapted to instrument→tested on instrument). The purpose of Experiments 2a and 2b were to gauge whether adaptation occurs similarly for different modalities (for voice and for instrumental sounds) by way of creating adaptation to a voice sound when testing on a voice sound (as in Experiment 2a). Also, the baseline conditions of Experiments 2a and 2b were used as stimulus verification. At step 1, sounds showed a lower averaged judgment score closer to anger with a score near 0, and at step 7 sounds received a higher averaged judgment score near 1, see Figure 6. This assured that sounds were initially representative of anger and fear, prior to adaptation.

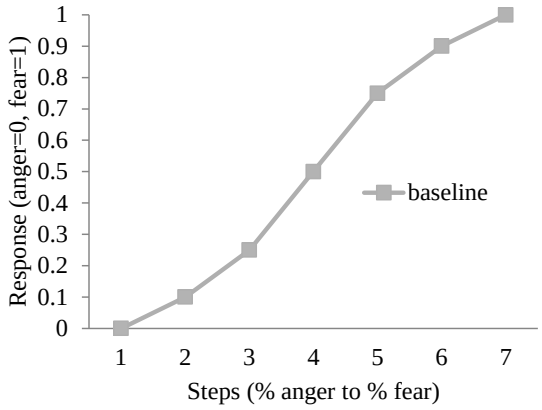


Figure 6. Example of the baseline phase for judgments of test sounds. The y-axis represents proportion of anger from participant's judgments of the morphed musical sounds, where 0 is the most angry and 1 is the least angry. The x-axis represents the morphed continuum for musical sounds where step 1 is the most angry and step 7 is the least angry.

In Experiment 2c, participants first heard voice sounds from the MAV in the exposure phase and in the test sound were asked to judge if an instrumental sound was angry or fearful (adapted to voice → tested on instrument). Experiment 2d was the opposite of Experiment 2c, where participants first heard an instrumental sound at exposure and a voice sound at test (adapted to instrument → tested on voice). See Figure 7 for a diagram of the experiment procedure. The purpose of Experiments 2c and 2d was to test for cross-modal adaptation aftereffects.

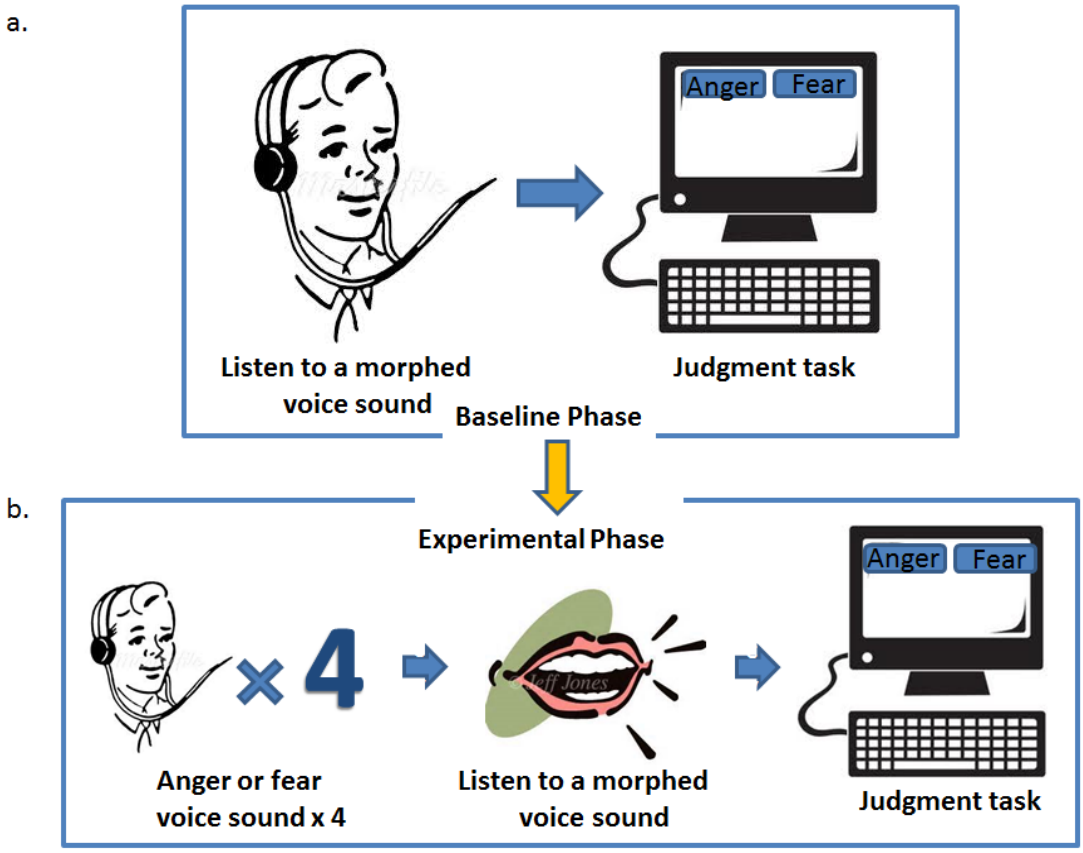


Figure 7. A schematic illustration of the baseline phase (a) and experimental phase (b) for Experiments 2a-2d. This illustration best depicts Experiment 2a with voice sounds; however, the procedure is the same for all experiments.

If emotion processing for these two types of sound make use of shared neural mechanisms, and if emotion processing in the two domains is related in terms of their motivational characteristics (Frühholz & Grandjean, 2013), one would predict that prolonged exposure to voice sounds (e.g., angry voice) should result in after effects (either adaptation or sensitization) in the processing of instrumental sounds and vice-versa.

3.2.2. Method

3.2.2.1. Participants. Twenty undergraduates participated in Experiment 2a (14 female, mean age = 19.1, SD = 1.35; 5 male, mean age = 20.6, SD = 3.71), (adapt to voice, test on voice) and 21 undergraduates took part in Experiment 2b (14 female, mean age = 19.57, SD = 2.06; 7 male, mean age = 18.57, SD = 1.51) (adapt to instrument, test on instrument). Thirty-six undergraduate students participated in Experiment 2c (adapt to voice, test on instrument) (19 female, mean age = 18.7, SD = 0.82; 17 male, mean age = 19.7, SD = 2.02). Fifty-two undergraduate students took part in Experiment 2d (adapt to instrument, test on voice) (24 female, mean age = 18.96, SD = 0.91; 28 male, mean age = 19.32, SD = 1.09). All participants reported normal hearing and received course credit.

3.2.2.2. Materials. For the instrumental sounds used in the baseline and experimental test phases, stimuli were created from instrumental recordings taken from two classes of musical instruments, brass and woodwind. Selected instruments were the French horn, baritone, saxophone, and flute, recorded at 440Hz. Instrumentalists from which the sounds were recorded were directed to play both an angry and a fearful sound for each instrument. From these recordings angry to fearful continua were created from each instrument in seven steps that corresponded to 5/95%, 20/80%, 35/65%, 50/50%, 65/35%, 80/20%, and 95/5% anger/fear. For the voice sounds used in the baseline and experimental test phases, stimuli were from two female and two male voices, taken from the Montreal Affective Voices (MAV, Belin, Fillion-Bilodeau & Gosselin, 2008). The MAV were designed as an auditory equivalent of the affective faces by Ekman and

Friesen (1986); these are nonverbal affect bursts that correspond to anger, disgust, fear, pain, sadness, surprise happiness and pleasure. Analyses of the MAV show a mean rating of 68% for valence and arousal, which indicates high recognition accuracy. These stimuli have been used by Bestelmeyer (2010; 2014). To create the MAVs, actors were instructed to produce emotional interjections used the vowel /a/. For prolonged exposure sounds, voices from four identities were chosen, two male, and two female; each expressing anger and fear. Stimuli were normalized in energy and presented in stereo via JVC Flats stereo headphones. The program STRAIGHT (Kawahari & Matsui, 2003) was used to create the anger-fear morphed continua in MatlabR2007b (Mathworks, Inc.).

3.2.2.3. Procedure. The experiment consisted of two phases—a baseline phase without prior prolonged exposure sounds and an experimental phase with prior prolonged exposure sounds. In the baseline phase, subjects received 84 trials, with 2 blocks of trials, one for each voice (2 male and 2 female) or instrument class (2 brass and 2 woodwind) which was always given prior to the experimental phase. Each sound at each of the seven morph steps was repeated six times, leading to 84 trials per voice or instrument block, with a total of 168 trials. Within each block, sounds were presented randomly with an inter-stimulus interval of 2-3s. Following the baseline phase participants took part in the experimental phase where the trial structure consisted of one voice or instrument played four times followed by an ambiguous morph after a silent gap of 1 second. There were four adaptation blocks (2 emotion x 2 gender or instrument) and each of the seven test stimuli per identity was repeated six times leading to 84 trials per block with a total of 336 trials. Table 5 summarizes the structure of the baseline and test

phases of Experiment 2a and 2b.

Table 5. Stimuli used in the baseline and adaptation phases in Experiments 2a-2d

Experiment Phase	Baseline	Adaptation	
		Exposure	Test
Exp. 2a	**Voice sounds: anger-fear judgment**	**Voice sounds**	**Voice sounds: anger-fear judgment**
Exp. 2b	**Instrumental sounds: anger-fear judgment**	**Instrumental sounds**	**Instrumental sounds: anger-fear judgment**
Exp. 2c	**Instrumental sounds: anger-fear judgment**	**Voice sounds**	**Instrumental sounds: anger-fear judgment**
Exp. 2d	**Voice sounds: anger-fear judgment**	**Instrumental sounds**	**Voice sounds: anger-fear judgment**

3.2.2.4. Design. For all data analyses, data were averaged as a function of the seven morph steps, where each participant had an average emotion judgment score for each sound at each step. A one-way repeated measures ANOVA was applied to the averaged judgment data.

3.2.3. Results

3.2.3.1. Experiment 2a - Voice → Voice. Prolonged exposure to an angry voice in Experiment 2a showed that participant's consistently judged voice sounds at test as more fearful, demonstrating an adaptation aftereffect. A one-way repeated measures ANOVA on behavioral responses revealed a significant main effect for affective voice sounds when participants were tested on voice sounds, Figure 8, ($F(2, 44) = 10.10$, $MSE = .036$, $p < .001$, $\eta 2p = .32$).

To examine the direction of this effect, paired t–tests were run and indicated that there was a significant difference for the baseline and anger conditions, t (22) = 4.63, p < .001, d = 1.05, 95% CI_d [.43, 1.69], where participants judged sounds as more fearful when exposed to anger (M = .61, SD = .09) relative to baseline (M = .52, SD = .07). A significant difference was also present for the anger versus fear conditions, t(22) = 3.06, p<.01, d = .40, 95% CI_d [.19, 1.00]. Participants judged sounds as more fearful when exposed to anger (M = .61, SD = .09) and more angry when exposed to fear (M = .56, SD = .09). The baseline versus fear condition was not significant.

a.

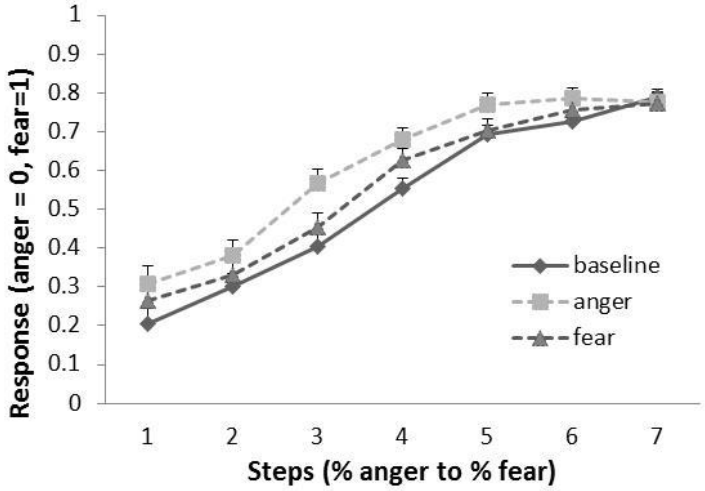
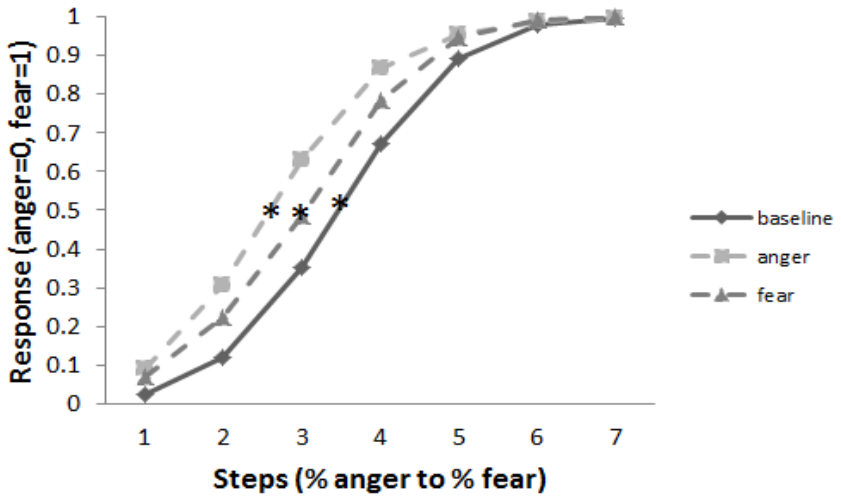


Figure 8. Behavioral results for prolonged exposure to voice sounds when tested on voice sounds (a). The grand average of all participants is displayed. Psychophysical function for the grand average of the three experimental conditions: baseline (solid), anger (light dashed) and fear (dark dashed). The points of subjective equality (PSE) values are denoted with a star (b).

Figure 8 continued.

b.



To further explore the direction of the effect, data were averaged as a function of the seven morph steps and a psychophysical curve (the hyperbolic tangent function) was fitted to the mean data for each adaptor type (baseline, anger and fear). Good fits were obtained for all three conditions; baseline (R^2 = .97), anger (R^2 = .99), and fear (R^2 = .98). The point of inflection of the function (point of subjective equality—PSE) was computed for all curves (baseline, anger and fear) as illustrated with an asterisk in Figure 8b. The point of inflection refers to the point on the test continuum where the instrument at test was equally likely to be labelled as angry or fearful.

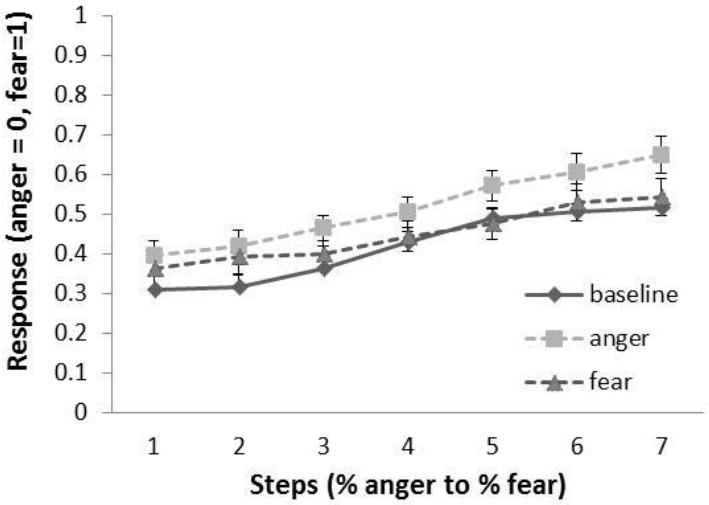
A one-way repeated measures ANOVA on inflection (PSE) values also revealed a significant main effect of adaptation to affective voices ($F(2, 44) = 7.12$, $MSE = .529$, $p < .01$, $\eta^2_p = .25$). Exploring the main effects with t-tests show that the PSE as a result of adaptation to anger was significantly smaller ($M = 2.65$, $SD = .97$) than the baseline

condition (M = 3.45, SD = .88), (t (22) = 3.35, p <.01), again showing that prolonged exposure to an angry voice produces adaptation. Additionally, fear was also significantly lower (M = 2.99, SD =2.13) than the baseline condition (M = 3.45 SD = .88), t (22) = 2.32, p <.05, again showing that adaptation occurs when participants were exposed to a fearful voice.

3.2.3.2. Experiment 2b - Instrument → Instrument. Similar to Experiment 2a, prolonged exposure to an angry sound results in adaptation to angry, but not fearful sounds. Experiment 2b revealed an adaptation effect for instrumental, rather than vocal sounds, showing the same effect in a different modality.

A one-way repeated measures ANOVA on behavioral responses revealed a significant main effect for affective instrumental sounds when participants were tested on instrumental sounds, Figure 9, (F (2, 38) = 3.81, MSE = .019, p < .001, η^2_p = .17). Planned t–tests indicate that participants exposed to angry instrumental sounds judged instrumental test sounds as more fearful (M =.52, SD = .16) compared to the baseline condition (M = .41, SD = .07); t (19) = 2.52, p < .05, d = .80, 95% CI_d [.13, 1.45]. There was no significant difference between the baseline and fear conditions or the anger versus fear conditions.

a.



b.

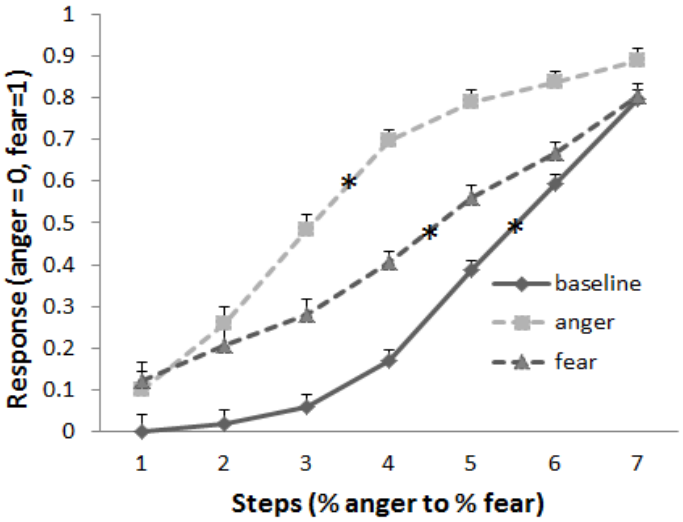


Figure 9. Behavioral results for prolonged exposure to instruments when tested on instrumental sounds (a). The grand average of all participants is displayed. Psychophysical function for the grand average of the three experimental conditions: baseline (solid), anger (light dashed) and fear (dark dashed). The PSE values are denoted with an asterisk (b).

The data were fitted with a psychophysical curve (the hyperbolic tangent function) where good fits were obtained for all three conditions; baseline (R^2 = .99), anger (R^2 = .95), and fear (R^2 = .96) (Figure 9b). A one-way repeated measures ANOVA on PSE values revealed a significant main effect of adaptation to affective instrument sounds ($F(2, 44) = 7.65$, $MSE = 2.811$, $p < .001$, $\eta^2_p = .26$). Planned t-tests showed that the PSE as a result of adaptation to anger was significantly smaller ($M = 3.45$, $SD = 2.13$) than the baseline condition ($M = 5.53$, $SD = 1.37$), ($t(22) = 3.701$, $p < .001$). In addition, anger was also significantly smaller ($M = 3.45$, $SD = 2.13$) than fear ($M = 4.51$, $SD = 2.45$), $t(22) = 2.30$, $p < .05$. These results suggest that prolonged exposure to an angry vocalization results in adaptation, after fitting the data to a psychophysical curve.

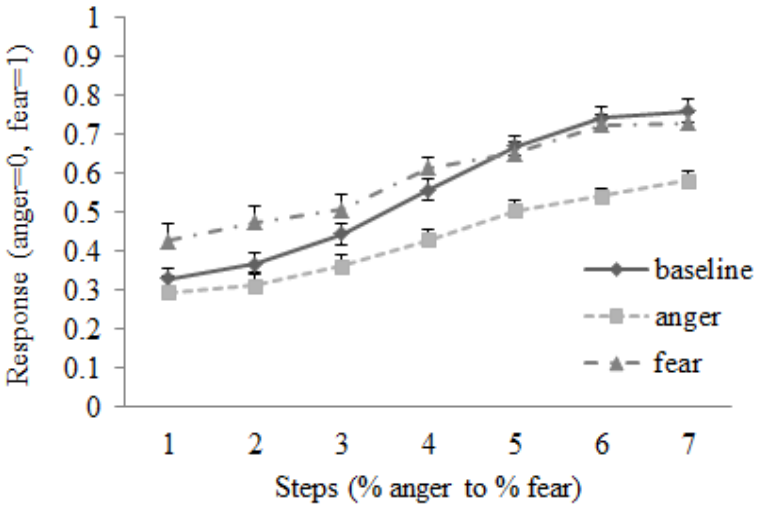
3.2.3.3. Experiment 2c - Voice → Instrument. Experiments 2a and 2b served as a stimulus validation to show that adaptation can occur in different modalities (voice and instrument). In Experiment 2c and 2d, I investigated the relationship between voice and instrumental sounds for cross-modal adaptation effects. Cross-modal effects were found when participants were exposed to anger, however, this resulted in sensitization where participants judged an instrumental test sound as more angry after prolonged exposure to an angry voice; however, there was no effect when participants were exposed to a fearful voice.

A one-way repeated measures ANOVA on behavioral responses revealed a significant main effect for affective voice sounds when participants were tested on instrumental sounds, Figure 10, ($F(2, 70) = 21.71$, $MSE = .070$, $p < .001$, $\eta^2_p = .38$). Planned t–tests indicate that there was a significant difference for the baseline and anger

conditions, t (35) = 4.61, $p < .001$, d = .91, 95% CI_d [.41, 1.40], where participants judged sounds as angrier after exposure to anger (M =.43, SD = .14), relative to baseline (M = .55, SD = .12). A significant difference was also present for the anger versus fear conditions, t(35) = 6.25, p<.001, d = 1.02, 95% CI_d [.52, 1.52]. Participants judged sounds as more fearful when exposed to fear (M = .59, SD = .17), relative to anger (M =.43, SD = .14). The baseline versus fear conditions was not significant.

As in the previous experiments, a psychophysical curve (the hyperbolic tangent function) was fitted to the mean data for each adaptor type (baseline, anger and fear) and good fits were obtained for all three conditions; baseline (R^2 = .76), anger (R^2 = .74), and fear (R^2 = .77), the PSEs are illustrated with an asterisk in Figure 10b. A one-way repeated measures ANOVA on PSE values showed a significant main effect of adaptation to affective voices (F(2, 68) = 17.41, MSE = .07, $p < .001$, η^2_p = .34). Planned t-tests show that the PSE as a result of adaptation to anger was significantly larger (M = 4.39, SD = 2.13) than the baseline condition (M = 3.31, SD = 1.41), (t(35) = 3.11, p <.05), supporting previous results that adaptation to an angry voice causes sensitization. In addition, anger was also rated significantly higher (M = 4.39, SD = 2.13) than fear (M = 2.69, SD = 2.10), t(35) = 6.41, p <.05.

a.



b.

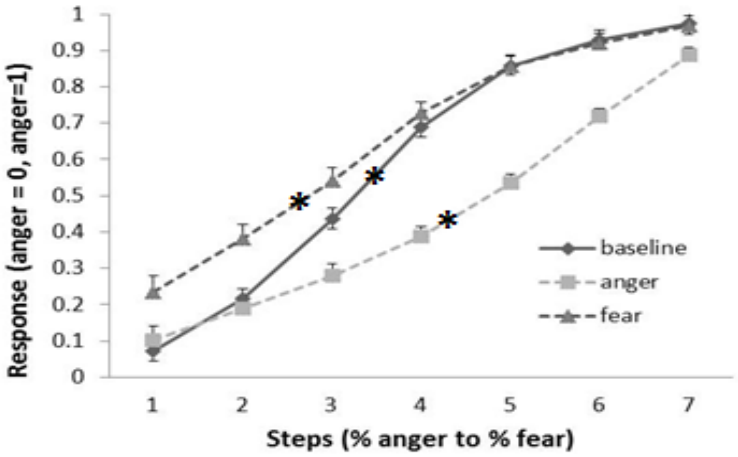
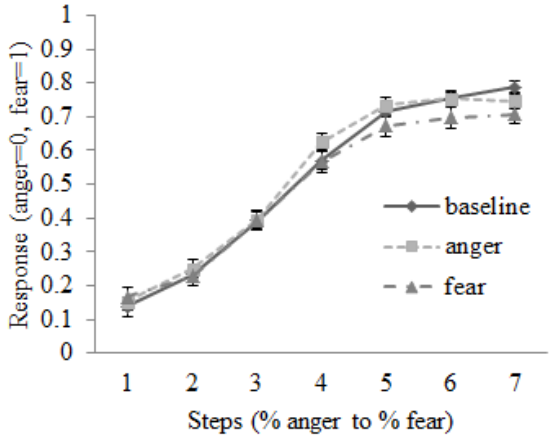


Figure 10. Behavioral results for prolonged exposure to voice sounds when tested on instrumental sounds (a). The grand average of all participants is displayed. Psychophysical function for the grand average of the three experimental conditions: baseline (solid), anger (light dashed) and fear (dark dashed). PSE values are illustrated with an asterisk (b).

3.2.3.4. Experiment 2d - Instrument → Voice. In contrast to the adaptation aftereffects in Experiments 2a and 2b, or the sensitization effect in Experiment 2c, there was no indication of adaptation or sensitization when participants were exposed to angry or fearful to instrumental sounds and tested on voice sounds, $F(2, 102) = 1.53$, $MSE = .065$, $p = .221$, $\eta^2_p = .029$, (Figure 11).

a.



b.

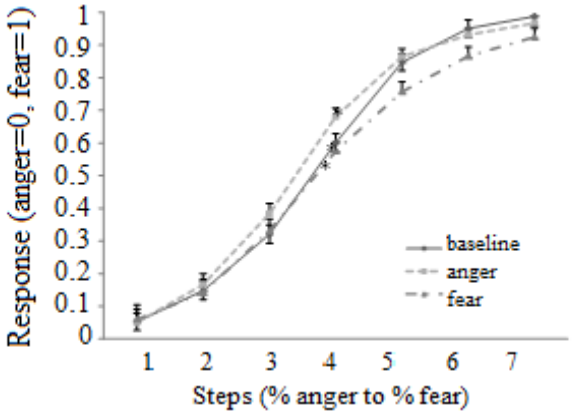


Figure 11. Behavioral results for prolonged exposure to instrumental sounds when tested on voice sounds (a). The grand average of all participants is displayed. Psychophysical function for the grand average of the three experimental conditions: baseline (solid), anger (light dashed) and fear (dark dashed) (b).

3.2.4. Discussion

The purpose of Experiments 2a-2d was to identify the extent to which emotion processing for voice and instrumental sounds could cross modalities and whether a common mechanism exists for emotion processing. Employing an adaptation framework modeled after Bestelmeyer et al. (2010; 2014), participants in Experiment 2a were exposed multiple times to an angry or fearful voice and judged whether a voice sound at test (on a morphed anger-fear continuum) was angry or fearful. Experiment 2b was similar except that participants judged whether an instrumental sound was angry or fearful after prolonged exposure to an angry or fearful instrumental sound. Experiments 2c and 2d tested for cross-modal aftereffects where in Experiment 2c participants were exposed multiple times to an angry or fearful voice sound and judged whether an instrumental test sound (on a morphed anger-fear continuum) was angry or fearful. Experiment 2d was the opposite of Experiment 2c where participants were exposed to an angry or fearful instrument sound and tested on a voice sound.

Results indicated that in Experiment 2a, exposure to angry voices made voice stimuli sound more fearful and less angry. Experiment 2b showed that participants judged instrumental sounds as more fearful when adapted to an angry sound and similar to Experiment 2a, showed no effect when adapted to fear. Experiment 2c demonstrated that exposure to angry voices made instrumental stimuli sound angrier and less fearful (sensitization), while exposure to fearful voices had no effect. Results from Experiment 2d showed no effect when participants were exposed to an angry or fearful instrumental sound. Overall, when exposed to angry voice sounds, listener's showed a marked

increase in fear responses. This indicates that affective voice sounds have an effect on the emotion perception of affective instrumental sounds. This result was not present for exposure to fearful voices or for repeated exposure to affective instrumental sounds.

The results from Experiments 2a and 2b (voice → voice and instrument → instrument) support previous research indicating that adaptation can take place in more than one modality (see Bestelmeyer et al., 2014). When participants were tested across modalities (e.g., prolonged exposure to voice and tested on instrumental sounds) there was a sensitization effect only for adaptation to angry sounds and no effect for adaptation to fearful sounds. This finding may reflect the difference in the underlying motivational salience (approach versus avoidance) for the emotions anger and fear. This indicates the possibility of a sub-mechanism used for processing different types of emotions. To better understand how this result could generalize to the domains of speech and music, it is necessary to use stimuli that better represent speech and music.

3.3. Music and speech

Similar to Experiments 2a-2d, the following studies used the same paradigm to directly compare the effect of anger and fear adaptation on emotion judgments for both musical (3 note sounds) and vocal sounds (2 phoneme vocal sounds). The domain of speech is represented by "speech–like" vocal sounds created from recordings of voices using the phonemes gi/go, wo/wo, de/de, or te/te. Musical sound stimuli represent the domain of music and are recordings of instrumental tones combined to create 3 note musical sounds. The study of comparing the domains of speech and music enables us to search for the hidden associations that can merge different phenomena (Patel, 2009) and

answer questions such as, what is the main link among emotion, music and non-linguistic speech.

3.3.1. Overview of experiments: 3a – vocal sound → vocal sound, 3b – musical sound → musical sound, 3c - vocal sound → musical sound and 3d - musical sound → vocal sound

Similar to Experiments 2a and 2b, Experiments 3a and 3b tested the validity of the vocal sound and musical sound stimuli. In Experiment 3a, participants were adapted to an angry or fearful vocal sound and tested on a morphed continuum of vocal sounds. In Experiment 3b participants were adapted to an angry or fearful musical sound (three note sound) and tested on a musical sound (three note sound). Experiments 3c and 3d examined if cross-modal aftereffects were present when adapting to an angry or fearful musical or vocal sound when tested on the opposite sound (vocal or musical sound, respectively), see Table 6. In addition, Experiments 3c and 3d further examined the difference found between anger and fear in Experiments 2c and 2d in terms of their motivational salience—approach and avoidance. Approach is associated with positive feelings, and avoidance with negative feelings (Cacioppo, Gardner & Berntson, 1999; Lang, 1995; Russell & Carroll, 1999; Watson, Wiese, Vaidya, & Tellegen, 1999); however, anger serves as a confound—anger is associated with approach but coupled with negative feelings (Eder et al., 2013; Harmon-Jones, Harmon-Jones, & Price, 2013; Harmon-Jones, 2003).This confound potentially motivates the difference in emotion perception between anger and fear.

The procedure for all experiments was similar to Experiments 2a-2d with a few key exceptions. In the baseline phase subjects heard a sound from the morphed test continuum that was either a vocal or musical sound (see Table 6) and judged if the sound was angry or fearful. In the experimental phase participants heard an angry or fearful vocal sound four times to elicit adaptation. Participants then heard a test sound from a morphed continuum ranging from anger to fear and judged whether the sound at test was angry or fearful. The impact of adaptation was analyzed by examining whether angry or fearful sounds had an effect on participants' anger-fear judgments for musical, vocal, or both types of sounds (cross-modal).

Table 6. Stimuli used in the baseline and adaptation phases of Experiments 3a-3d.

Experiment Phase	**Baseline**	**Adaptation**	
		Exposure	Test
Exp. 3a	**Vocal sounds: anger-fear judgment**	**Vocal sounds**	**Vocal sounds: anger-fear judgment**
Exp. 3b	**Musical sounds: anger-fear judgment**	**Musical sounds**	**Musical sounds: anger-fear judgment**
Exp. 3c	**Musical sounds: anger-fear judgment**	**Vocal sounds**	**Musical sounds: anger-fear judgment**
Exp. 3d	**Vocal sounds: anger-fear judgment**	**Musical sounds**	**Vocal sounds: anger-fear judgment**

3.3.2. Method

3.3.2.1. Participants. Seventeen undergraduate students took part in Experiment 3a (adapted to vocal sound → tested on vocal sound) (8 female, mean age = 19.00, SD = 0.53; 9 male, mean age = 19.67, SD = 1.41); 18 undergraduate students took part in

Experiment 3b (adapt to musical sound, test on musical sound) (10 female, mean age = 18.40, SD = 0.70; 8 male, mean age = 20.00, SD = 3.30); 20 undergraduate students participated in Experiment 3c (adapted to vocal sound → tested on musical sound) (12 female, mean age = 19, SD =1.12; 8 male, mean age = 20.4, SD = 2.56); and 20 undergraduate students participated in Experiment 3d (adapted to musical sound → tested on vocal sound) (12 female, mean age = 19.20, SD = 1.94; 8 male, mean age = 20.37, SD = 2.77). All participants reported normal hearing and received course credit.

3.3.2.2. Materials. Musical sound stimuli were 168 sounds, each of which lasted between 1.5 and 3 seconds. These musical sounds were modifications of instrumental sounds employed in Bowman and Yamauchi (in press), where individual instrumental sounds were created from recordings of two classes of musical instruments, brass and woodwind, performed by members of the U.S. 395th Army band. Selected instruments were the French horn, baritone, saxophone, and flute, recorded at 440Hz. Instrumentalists from which the sounds were recorded were directed to play both an angry and a fearful sound for each instrument. To create the three note musical sound stimuli, three angry or fearful instrumental sounds were combined to create a three note musical sound. From these three note musical sound stimuli, angry to fearful continua were created from each sound in seven steps that corresponded to 5/95%, 20/80%, 35/65%, 50/50%, 65/35%, 80/20%, and 95/5% anger/fear. For the prolonged exposure sounds used in the experimental phase, the original angry (0/100%) and fearful (100/0%) musical sounds for each instrument were used as adaptors. All stimuli were normalized in energy and presented in stereo via JVC Flats stereo headphones. As in Experiments

2a-2d, the program STRAIGHT (Kawahara & Matsui, 2003) was used to create the anger/fear morphs.

Vocal sound stimuli consisted of 168 pseudo speech sounds recorded by four actors and modified after those used in Klinge, Röder, & Büchel (2010). Angry to fearful continua were created separately for each voice identity (male or female), in seven steps that corresponded to 5/95%, 20/80%, 35/65%, 50/50%, 65/35%, 80/20% and 95/5% anger/fear in the same manner used to create musical sounds.

3.3.2.3. Procedure. The procedure was similar to Experiments 2a-2d and was the same for all Experiments 3a-3d, with exception to the sounds presented. Experiments consisted of two main parts, a baseline phase without prior prolonged exposure and an experimental phase with prolonged exposure to an anger or fear sound, see Figure 12.

The baseline phase consisted of 84 trials in two blocks, one for male sounds and one for female sounds (vocal sounds, Experiments 3a and 3d) or one for woodwind and one for brass (musical sounds, Experiments 3b and 3c), given prior to the adaptation task. In the baseline phase participants received 168 sounds one at a time and judged whether each sound was angry or fearful. The sound of each identity (gender or instrument type; woodwind or brass) at each of the seven morph steps was repeated six times, resulting in 84 baseline trials per block with a total of 168 trials (4 voices/instruments x 7 anger-fear morphed steps x 6 times = 168 trials). Within each block sounds were presented randomly with an inter-stimulus interval of 2 seconds. In each trial, participants heard a sound (vocal or musical sound) from one of the seven

vocal or musical sound morphed steps and were asked judged whether the sound was angry or fearful (i.e., anger-fear judgment task).

The experimental phase was similar to the baseline phase except that vocal or musical sounds presented in the baseline phase except that sounds at test were preceded by either an angry or fearful vocal or musical sound, yielding 336 trials; 2 (angry or fearful) vocal or musical sounds x 4 voices x 7 anger-fear morphed steps x 6 times = 336 trials. Participants were tested on a different identity than the one they were adapted to (e.g., in Experiment 3a vocal sound-vocal sound, they were adapted to a female, and tested on male), to avoid low-level adaptation to factors such as voice identity.

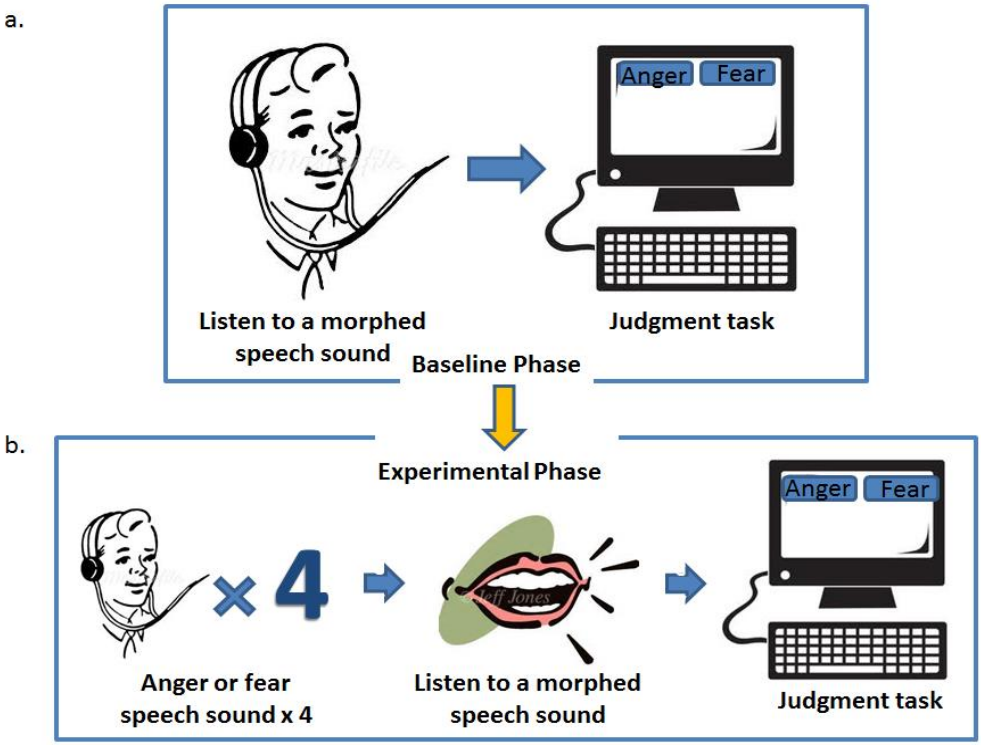


Figure 12. A schematic illustration of the baseline phase (a) and experimental phase (b) for Experiments 3a-3d. This illustration best depicts Experiment 3a with vocal sounds; however, the procedure was the same for all experiments.

3.3.2.4. Design. The dependent variable was the proportion of trials that participants judged stimulus sounds as angry or fearful and the independent variable was the prolonged exposure condition (baseline, anger, or fear). In the baseline condition, participants received no prior sound stimuli; in the angry stimuli exposure condition, angry vocal or musical sounds were given prior to the test sound and the anger-fear judgment task; in the fearful stimuli exposure condition, fearful vocal or musical sounds were presented before the test sound and anger-fear judgment task.

3.3.2.5. Analyses. Analyses were the same as used in Experiments 2a-2d where data were averaged as a function of the seven morph steps. The experiments used a within-subjects design and a one-way repeated measures ANOVA was applied to assess differences between the baseline, anger and fear conditions.

3.3.3. Results

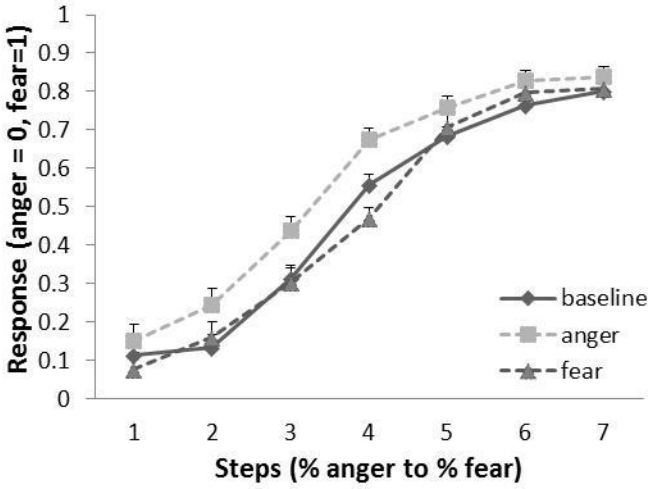
3.3.3.1. Experiment 3a – Vocal sound→ Vocal sound. Prolonged exposure to angry vocal sounds revealed that participants judged vocal sounds at test as more fearful, showing an adaptation effect similar to Experiment 2a (voice-voice). A one-way repeated measures ANOVA revealed a significant main effect for affective vocal sounds when participants were tested on vocal sounds, ($F(2, 32) = 4.19$, $MSE = .055$, $p < .05$, $\eta^2_p = .21$), Figure 13.

Paired t–tests show a significant difference between the baseline and anger conditions, $t(16) = 2.21$, $p < .05$, $d = .82$, 95% CI_d [.08, 1.53], where participants judged sounds as more fearful when exposed to anger ($M = .56$, $SD = .11$) relative to baseline ($M = .48$, $SD = .08$). A significant difference was also present for the anger versus fear

conditions, $t(16) = 5.18$, $p<.001$, $d = .72$, 95% CI_d [.05, 1.43]. Participants judged sounds as more fearful when exposed to anger (M =.56, SD = .11) and more angry when exposed to fear (M = .47, SD = .12). The baseline versus fear condition was not significant.

As in Experiments 2a-2d, data were averaged as a function of the seven morph steps and a psychophysical curve (the hyperbolic tangent function) was fitted to the mean data for each adaptor type (baseline, anger and fear). Good fits were obtained for all three conditions; baseline (R^2 = .98), anger (R^2 = .99), and fear (R^2 = .98) and the point of inflection of the function was computed for all curves, as illustrated with an asterisk in Figure 13b. A one-way repeated measures ANOVA on inflection values revealed that there was no main effect of adaptation to affective vocal sounds, ($F(2, 32) = 2.69$, MSE = 1.18, $p > .05$, $\eta^2_p = .14$).

a.



b.

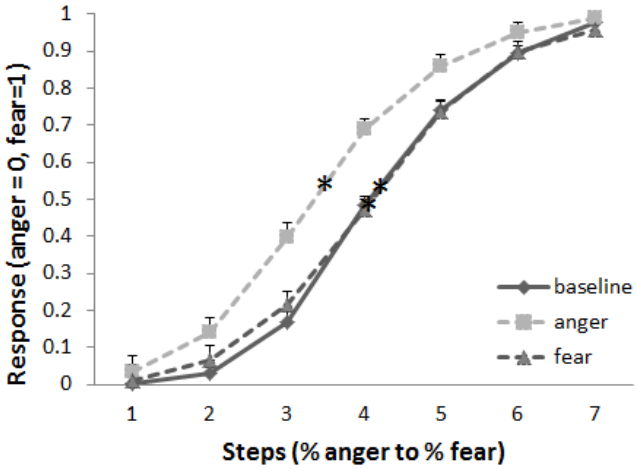


Figure 13. Behavioral results for prolonged exposure to vocal sounds when tested on vocal sounds (a). The grand average of all participants is displayed. Psychophysical function for the grand average of the three experimental conditions: baseline (solid), anger (light dashed) and fear (dark dashed). The PSE values are denoted with an asterisk (b).

3.3.3.2. Experiment 3b – Musical sounds → Musical sounds. Similar to Experiment 3a, Experiment 3b functions as a stimulus validation for musical stimuli.

Prolonged exposure to angry musical sounds in Experiment 2b showed that participant's consistently judged musical sounds as more fearful, demonstrating an adaptation effect. Similarly, when participants were exposed to a fearful musical sound, they consistently judged musical sounds at test as more angry Figure 14, (F (2, 30) = 18.10, MSE = .027, p < .001, η^2_p = .55).

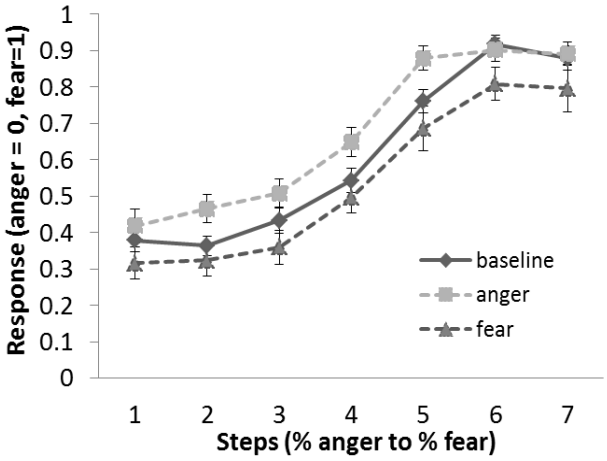
Paired t–tests indicated that there was a significant difference for the baseline and anger conditions, t (15) = 3.35, p < .01, d = .65, 95% CI_d [.06, 1.39], where participants judged sounds as more fearful when exposed to anger (M =.67, SD = .11), relative to baseline (M = .61, SD = .08). An adaptation effect was also present for the baseline and fear conditions, t(16) = 2.61, p<.01, d = .60, 95% CI_d [.13, 1.33]. Participants judged sounds as more angry when exposed to fear (M =.54, SD = .15) compared to baseline (M = .61, SD = .08). A difference was also present for the anger and fear conditions, t (15) = 6.76, p < .001, d = 1.02, 95% CI_d [.26, 1.79].

Fitting the data to a psychophysical curve (the hyperbolic tangent function), good fits were obtained for all three conditions; baseline (R^2 = .99), anger (R^2 = .99), and fear (R^2 = .99). The PSEs for each condition are illustrated with an asterisk in Figure 14b.

A one-way repeated measures ANOVA on inflection values revealed a significant main effect of adaptation to affective musical sounds, (F(2, 30) = 8.76, MSE = .87, p < .001, η^2_p = .37). Follow up t-tests showed that the PSE as a result of adaptation to anger was significantly smaller (M = 2.60, SD = 1.22) compared to baseline (M = 3.35, SD = .91), t(15) = 2.70, p < .01). Additionally, the PSE as a result of adaptation to fear was significantly larger (M = 3.98, SD = 1.71) compared to baseline, t(15) = 2.10,

p<.05). These results are consistent with those prior to the curve fitting and demonstrate an effect of adaptation when participants were exposed to anger and when participants were exposed to fear.

a.



b.

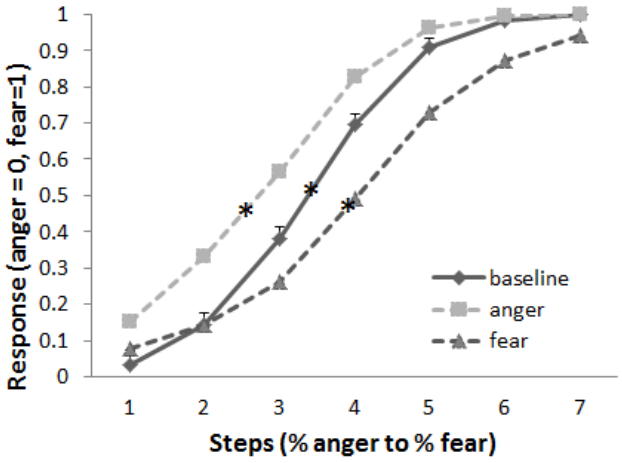


Figure 14. Behavioral results for prolonged exposure to musical sounds when tested on musical sounds (a). The grand average of all participants is displayed. Psychophysical function for the grand average of the three experimental conditions: baseline (solid), anger (light dashed) and fear (dark dashed). The PSE values are represented by an asterisk (b).

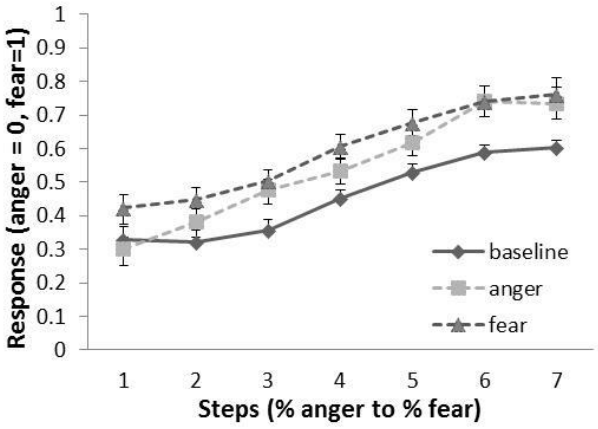
3.3.3.3. Experiment 3c – Vocal sounds → Musical sounds. Prolonged exposure to angry vocal sounds demonstrated an adaptation effect where participants' judged musical sounds as more fearful. Similarly, when exposed to a fearful vocal sound, participants also judged musical sounds at test as more fearful, demonstrating a sensitization effect; Figure 15, (F (2, 38) = 10.38, MSE = .068, $p < .001$, $\eta^2_p = .35$).

Paired t–tests indicate that there was a significant difference for the baseline and anger conditions, t (19) = 2.94, $p < .01$, d = .69, 95% CI_d [.03, 1.34], where participants judged sounds as more fearful when exposed to anger (M =. 54, SD = .14) relative to baseline (M = .45, SD = .10). A sensitization effect was found when participants were exposed to fear t(19) = 4.43, p<.001, d = 1.03, 95% CI_d [.35, 1.71] where sounds were judged as more fearful when exposed to fear (M =.59, SD = .16) relative to baseline (M = .45, SD = .10). A difference was not present for the anger and fear conditions.

Good fits were obtained for the data after fitting to the hyperbolic tangent function; baseline (R^2 = .98), anger (R^2 = .92), and fear (R^2 = .98), the PSEs are illustrated with an asterisk in Figure 15b. A one-way repeated measures ANOVA on PSE values revealed a significant main effect of adaptation to affective vocal sounds when tested on musical sounds, (F(2, 38) =4.03, MSE = 2.23, $p < .05$, $\eta^2_p = .18$). Follow up t-tests showed that the PSE as a result of adaptation to fear was significantly smaller (M = 2.93, SD = 2.01) compared to baseline (M = 4.21, SD = 1.71), t(19) = 2.77, $p <$ .01). There was no difference for baseline compared to adaptation to fear or for anger compared to fear. These results are in agreement with those prior to curve fitting that show an effect of adaptation when participants are exposed to an angry vocal sound and

tested on a musical sound, but do not show the same sensitization effect. This could indicate that the effect of sensitization is not as strong as adaptation.

a.



b.

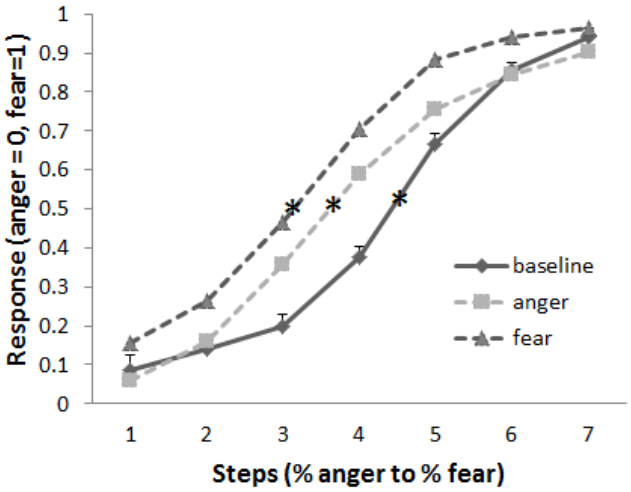


Figure 15. Behavioral results for prolonged exposure to vocal sounds when tested on musical sounds (a). The grand average of all participants is displayed. Psychophysical function for the grand average of the three experimental conditions: baseline (solid), anger (light dashed) and fear (dark dashed). PSE values are denoted with an asterisk (b).

3.3.3.4. Experiment 3d – Musical sound → Vocal sound. Prolonged exposure to angry musical sounds did not cause participants to judge sounds as more angry or fearful at test. Similarly, prolonged exposure to fearful musical sounds did not cause participants to judge sounds as more angry or fearful at test (F (2, 38) = 2.92, MSE = .028, p> .05, η^2_p = .13), Figure 16.

a.

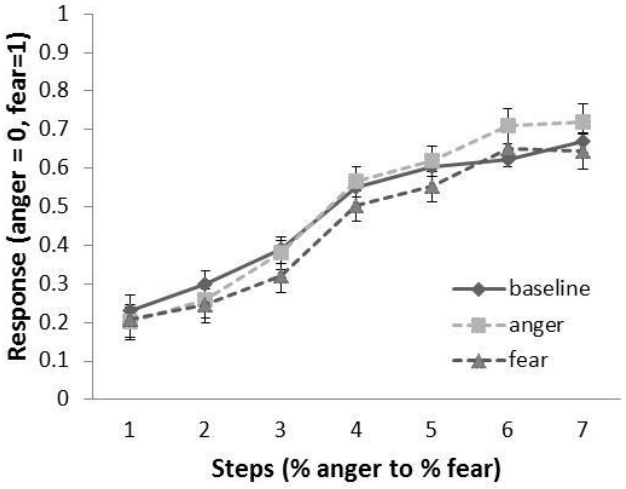
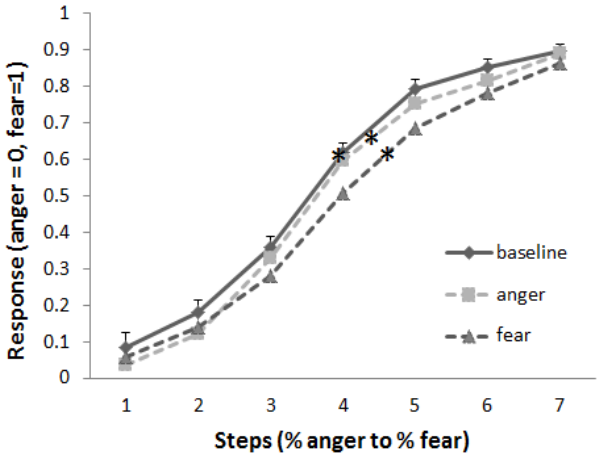


Figure 16. Behavioral results for prolonged exposure to musical sounds when tested on vocal sounds (a). The grand average of all participants is displayed. Psychophysical function for the grand average of the three experimental conditions: baseline (solid), anger (light dashed) and fear (dark dashed) (b).

Figure 16 continued.

b.



3.3.4. Discussion

The purpose of these studies was to further investigate whether vocal and musical sounds use a common emotion processing mechanism and examine if any cross-modal effects of adaptation occurred, using stimuli that more closely resembled speech and music. Results from Experiment 3a demonstrated an adaptation effect where exposure to angry vocal sounds made vocal stimuli sound more fearful. Experiment 3b also showed an adaptation effect for angry and for fearful musical sounds when tested on musical sounds. Experiment 3c similarly revealed adaptation to angry vocal sounds, where participants judged musical sounds as more fearful when adapted to anger and a sensitization effect where participants judged musical sounds as more fearful when adapted to a fearful vocal sound. There were no adaptation or sensitization effects present for Experiment 3d when exposed to musical sounds and tested on vocal sounds.

Similar to Experiments 2a-2d which found an adaptation effect when participants were exposed to and tested in the same modality (e.g., voice-voice), this set of experiments demonstrated that participants exposed to *angry* vocal sounds and tested on vocal sounds and exposed to angry musical sounds and tested musical sounds, judged sounds at test as more *fearful*. Results from Experiments 3c and 3d testing the cross-modal effects of emotion perception, showed adaptation when participants were exposed to an angry vocal sound and sensitization when exposed to a fearful vocal sound and tested on musical sounds; however, neither adaptation nor sensitization was found when participants were exposed to a musical sound and tested on a vocal sound. While both are negatively valenced emotions, these results provide evidence indicating a difference in processing the emotions anger and fear.

CHAPTER IV

DISCUSSION AND CONCLUSIONS

4.1. Summary

The purpose of these studies was to uncover the link between the domains of music and speech with regard to acoustic components of sound and gauge whether adaptation occurs and crosses over the speech and music domains, signifying a shared emotion processing mechanism. Results showed that there was (1) a link between vocal and instrumental sounds, where similar acoustic components were used for emotion perception and (2) that similar adaptation aftereffects occurred for the perception of angry voice, instrumental, vocal and musical sounds.

These results provide evidence that there are similar mechanisms at play for emotion perception in the speech and music domains (represented by voice, instrumental, vocal and musical sounds) and also that the nature of this relationship is more complex than a simple shared mechanism. Specifically, there is likely a unidirectional relationship where vocal sounds can encompass musical sounds but not vice-versa. In addition, anger and fear were perceived differently such that prolonged exposure to anger caused adaptation, but not prolonged exposure to fear (see Experiment 2b, instrument-instrument and 3c, vocal sound-musical sound). Anger and fear are both negatively valenced emotions; however, they have differing motivational aspects that potentially drive the difference in perception. These ideas have not previously been considered across the speech, music and emotion literature and will be outlined further in the discussion section.

4.2. Discussion

4.2.1. Acoustic components of speech and music

Previous emotion research in the domains of speech and music has examined the effect of emotion on vocal acoustics (Bachorowski & Owren, 2008); infant-directed speech (Byrd et al., 2011; Schachner & Hannon, 2011) and music (Coutinho, Deng, & Schuller, 2014) and many studies suggest that there are shared mechanisms for emotion perception, yet few studies have explored the link between speech and music.

Though studies of emotion in music and speech are predominately separate, they provide evidence for overlapping attributes in the two domains. For example, Eerola, Friberg & Bresin (2013), showed that acoustic features (mode, tempo, dynamics, articulation and timbre) contributed to the perception of emotion in music. Similarly, Byrd, Bowman and Yamauchi (2012) showed that acoustic features related to timbre explained emotion in infants' vocalizations (cooing and babbling). In contrast, an important finding of the present studies (Experiments 1a-1c) is the overlap of acoustic components used for emotion perception for both instrumental and voice sounds. This set of studies examined whether the same acoustic components could explain emotion perception in both the music and speech domains.

Studies using regression are helpful in uncovering features of sound that can explain emotion in music and speech, but they are limited. Regression is correlational and as such, is unsuitable to uncover the functional specificity underlying speech and music (e.g., whether the same or different neural mechanisms mediate emotion

processing in speech and music) (see Bestelmeyer et al., 2010 for exceptions and Juslin & Laukka 2003).

4.2.2. Directionality of emotion perception

A growing body of research has found support for the relationship between music and speech processing such that they share overlapping cognitive resources (Früholz, Trost, & Grandjean, 2014); however, less is known about the directionality of emotion perception in music and speech sounds. Levy, Granot and Bentin (2001) compared responses to voices and musical instruments using event related potentials (ERPs) and found evidence for a directional mechanism of emotion perception. Their results show a voice-specific response to sung voices and tones of musical instruments where mechanisms were more activated by voice-stimuli compared to non-vocal stimuli (Levy, Granot & Bentin, 2001; 2003; Belin, Fecteau & Bédard, 2004). This 'voice-specific' response is related to the salience of voice stimuli, reflecting the way attention is allocated, and suggests that emotion perception is mediated by vocalizations.

An important contribution of the adaptation studies in this dissertation support the aforementioned 'voice-specific' response (Levy, Granot & Bentin, 2001). Experiments 2a-2d demonstrated adaptation to angry voice and instrumental sounds when tested on voice and instrumental sounds (Experiments 2a and 2b, respectively) and cross-modal adaptation that only occurred when participants were exposed to an angry voice sounds and tested on an instrumental sounds (Experiment 2c). This same effect was found with vocal and musical stimuli where adaptation aftereffects occurred from vocal sound to vocal sound (Experiment 3a) and musical sound to musical sound

(Experiment 3b). More notably, this effect only occurred from vocal to musical sounds (Experiment 3c) not vice-versa, similar to Experiment 2c. These findings indicate that there is potentially a specific directionality for emotion perception in voice, instrumental, vocal and musical sounds; significantly, this unidirectional relationship reveals that mechanisms used for emotion processing may be shared from voice sounds to instrumental sounds and from vocal to musical sounds, but not vice-versa. This is in line with other studies that show a unidirectional auditory mechanism for speech and music perception. These studies, however, did not take into account whether participants had a strong background in musical experience or training, which could affect judgments of sounds. In addition, it is unclear whether adaptation aftereffects occurred due to adaptation to affect (anger and fear) or association occurred. These problems are addressed further in the limitations section.

4.2.3. Motivational salience and emotion perception

The results of these studies suggest that in addition to a unidirectional mechanism, there are potentially sub-mechanisms used for processing different emotions. An adaptation aftereffect was found for the cross-modal experiments (Experiments 2c and 2d and 3c and 3d) for the emotion anger and not fear where responses were either significantly decreased (adaptation) or increased (sensitization) when participants were repeatedly exposed to angry vocalizations (see Experiments 2a-2c and Experiments 3a-3c). These results indicate a difference in the way anger and fear are perceived. This difference is likely due to the adaptive value of the emotion anger compared to fear (Strauss et al., 2005).

In Bowman and Yamauchi (in press), when participants were adapted to an angry vocal sound, they judged a vocal sound at test as more angry. When exposed to a fearful vocal sound, however, participants did not judge an angry vocal sound as different. Because the motivational salience of sound plays an important role in the perception of emotion, this difference is potentially due to the adaptive value of emotion. In other words, rather than one mechanism processing these emotions altogether, they are likely processed in different channels according to their motivational salience (Strauss et al., 2005).

Aubé, Angulo-Perkins, Peretz, Concha and Armony (2014) addressed whether brain regions associated with processing the adaptive value of affective expressions were also employed by affective music. Using an event-related fMRI, responses to basic emotions (fear, sadness and happiness, as well as neutral) expressed through faces, nonlinguistic vocalizations and short, novel musical excerpts were compared. Results showed that responses in the amygdala to fearful music and vocalizations were correlated, revealing that the mechanisms used for emotion processing in music are shared with mechanisms that evolved for vocalizations (Strauss et al., 2005); though, this does not address whether emotion processing in music is mediated by emotion processing of voices.

Overall, within the music, speech and emotion literature, effects have been found to indicate the possibility of a directional mechanism for emotion perception and a sub-mechanism that processes categorical differences in emotion. Evidence from ERP studies show voice-specific responses where brain mechanisms are more activated by

vocal stimuli, and studies using fMRI have shown specific regions in the auditory cortex that elicit a greater response for vocal sounds, revealing a unidirectional mechanism that mediates emotion perception of vocal and instrumental sounds. Additionally, rating studies have shown that angry sounds are perceived as threatening, thereby increasing adaptive behaviors. Behavioral and fMRI studies of face perception demonstrate a sensitization to angry faces in the amygdala, which reflects a categorical difference for processing expressed emotion. Taken together, these studies and the studies in this dissertation support a unidirectional mechanism of emotion processing for vocal and musical sounds and supportive evidence for a sub-mechanism that is used to process categorical differences in sound, particularly anger and fear.

4.3. Limitations

As with previous studies joining the domains of emotion, speech and music, some limitations apply. First, effects of adaptation are difficult to interpret. It is unclear whether neural adaptation occurred, such that participants were adapted to an emotion (anger or fear), or whether participants were making judgments simply based on comparing sounds during exposure to sounds at test. Second, adaptation paradigms generally use a bottom-up approach where prior experiences of participants are not considered as impacting affective judgments.

Adaptation paradigms have been used in vision, emotion, and face perception but results are difficult to interpret because it is not clear if aftereffects are due to actual adaptation to emotion, or an association between adapting and test stimuli. Past adaptation research has questioned whether aftereffects found in face adaptation were

due to low-level (adaptation at retinal level) or high-level adaptation (adaptation in areas of the brain responsible for face processing) (Bestelmeyer et al., 2014). In Bestelmeyer et. al. (2010), this limitation was addressed by participants adapting to and testing on different voice identities. For instance, if adapting to a male voice, participants were tested on a female voice such that adaptation across vocal modulations was not likely due to low-level adaptation (e.g., to pitch of a voice), but instead high-level adaptation (e.g., to affect). Additionally, to combat this drawback it seems necessary to include physiological measures that help assure that adaptation is taking place. For example, physiological measures such as heart rate or skin conductance could indicate whether participants are simply responding to a sound stimulus (e.g., a fearful sound could produce higher skin conductance and heart rate) or whether participants are adapting to sounds during prolonged exposure.

A more inclusive adaptation paradigm needs to be formed that does not focus solely on sensory processes (bottom-up processing), but also includes prior experiences of participants that are likely to impact decision making (top-down processes). Research indicates that those with experience in music (reading music or playing an instrument) will be more proficient with speech related tasks (Juslin & Laukka, 2003). In these dissertation studies a unidirectional effect was found such that aftereffects occurred after prolonged exposure to vocal sounds when testing on musical sounds, but not vice versa. This could be interpreted that the participants in these studies did not have much musical experience such that they could not use information from music to make a decision about a vocal sound. To rule out this interpretation, a more diverse group of participants

needs to be used that includes musicians and those with musical experience. If similar aftereffects still occurred, it would be easier to say that there is a unidirectional effect such that speech-related sounds encompass musical sounds.

4.4. Future directions

Future studies should explore overlapping emotion processing that occurs in different types of sound stimuli such as the voice, music, and environmental sounds. This should include an expanded set of emotions for adaptation studies, rather than only anger and fear. A wider variety of stimuli could be employed that would increase the variability of participant's stimuli ratings. For example, including more instruments when creating musical stimuli, or using speech sounds that contain more speech information (e.g., the non-word /de/de/). It would be interesting to include natural sounds such as rainfall or a growling dog, compared to sounds that are produced by human action, such as the sound of a running bus engine. Future regression studies utilizing other types of sound, such as sounds from nature, may help to explain how affect is perceived and whether this perception does lie on a continuum, rather than completely separate scales. In addition, addressing this in future adaptation work could expand upon the idea of cross-modal adaptation in different domains and help to discover if there are shared emotion processing mechanisms.

Current research on speech and emotion has focused on processing of meaning through the semantic, lexical, conceptual, and propositional processing of language. However, music is also a means of communication and meaning also emerges from the interpretation of musical information (Koelsch, 2011). Sound symbolism is the idea in

linguistics, that there is a non-arbitrary relationship between the physical aspect of a speech signal and its meaning (Hinton, Nichols, & Ohala 1994; Ohala, 1994). Because musical sounds also communicate emotion and there is an overlap in processing for vocal and musical sounds, sound symbolism is likely evident in musical and other types of sounds as well. This is similar to embodied cognition in music, where the human body is a mediator between the mind and the physical environment. If this is the case that meaning in different forms is a mediated by the body, then there should be a strong relationship between the music and speech domains where a person can use a vocal sound, for example, to determine meaning in an instrumental or musical sound.

While there is a clear category boundary between anger-fear or anger-sadness continua for face and voice adaptation (Bestelmeyer et al., 2010), the boundary between these emotions may not be defined enough to encompass emotions shared between music and speech. In addition, the forced choice task paradigm may not allow for enough variety in responses to account for musical emotions in that there is not enough variability as compared to arousal and valence ratings of emotion. This research will encourage the building of a model for emotion perception in speech and music and further specify psychological mechanisms used for emotion processing.

REFERENCES

Agus, T. R., Suied, C., Thorpe, S. J., & Pressnitzer, D. (2012). Fast recognition of musical sounds based on timbre. *The Journal of the Acoustical Society of America*, *131*(5), 4124. http://doi.org/10.1121/1.3701865

Aubé, W., Angulo-Perkins, A., Peretz, I., Concha, L., & Armony, J. L. (2014). Fear across the senses: Brain responses to music, vocalizations and facial expressions. *Social Cognitive and Affective Neuroscience*, *10*, 399-407. http://doi.org/10.1093/scan/nsu067

Bachorowski, J.-A. (1999). Vocal expression and perception of emotion. *Current Directions in Psychological Science*, *8*(2), 53–57. http://doi.org/10.1111/1467-8721.00013

Bachorowski, J.-A., & Owren, M. J. (2008). *Handbook of emotions. New York* (Vol. 54). http://doi.org/10.2307/2076468

Bachorowski, J.-A., Smoski, M. J., & Owren, M. J. (2001). The acoustic features of human laughter. *The Journal of the Acoustical Society of America*, *110*(3), 1581. http://doi.org/10.1121/1.1391244

Balkwill, L., & Thompson, W. F. (1999). A cross-cultural investigation of the perception of cultural cues and emotion in music: Psychophysical and Cultural Cues. *Music Perception*, 43-64. http://doi.org/10.2307/40285811

Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psycholgy*, *70*(3), 614–636.

Bänziger, T., Patel, S., & Scherer, K. R. (2014). The role of perceived voice and speech characteristics in vocal emotion communication. *Journal of Nonverbal Behavior*, *38*(1), 31–52. http://doi.org/10.1007/s10919-013-0165-x

Behrens, G. (1993). the ability to identify emotional content of solo improvisations performed vocally and on three different instruments. *Psychology of Music*, *21(1)*, 20–33.

Belin, P., Fillion-Bilodeau, S., & Gosselin, F. (2008). The montreal affective voices: A validated set of nonverbal affect bursts for research on auditory affective processing. *Behavior Research Methods*, *40*(2), 531–539. http://doi.org/10.3758/BRM.40.2.531

Bergeson, T. R., & Trehub, S. E. (2007). Signature tunes in mothers' speech to infants.

Infant Behavior and Development, *30*(4), 648–654. http://doi.org/10.1016/j.infbeh.2007.03.003

Bestelmeyer, P. E. G., Jones, B. C., DeBruine, L. M., Little, a. C., & Welling, L. L. M. (2010). Face aftereffects demonstrate interdependent processing of expression and sex and of expression and race., *Visual Cosnition*, *18(2)*, 255-274. http://doi.org/10.1080/13506280802708024

Bestelmeyer, P. E. G., Maurage, P., Rouger, J., Latinus, M., & Belin, P. (2014). Adaptation to vocal expressions reveals multistep perception of auditory emotion. *The Journal of Neuroscience*, *34*(24), 8098–105. http://doi.org/10.1523/JNEUROSCI.4820-13.2014.

Bowman, C., & Yamauchi T. (in press). Perceiving categorical emotion in sound: The role of timbre. *Psychomusicology: Music, mind, and brain*.

Bregman, a S., Liao, C., & Levitan, R. (1990). Auditory grouping based on fundamental frequency and formant peak frequency. *Canadian Journal of Psychology*, *44*(3), 400–413. http://doi.org/10.1037/h0084255

Breiman, L. (University of C. (2001). Random forest. *Machine Learning*, *45*(5), 1–35. http://doi.org/10.1023/A:1010933404324

Brück, C., Kreifelts, B., & Wildgruber, D. (2012). From evolutionary roots to a broad spectrum of complex human emotions: Future research perspectives in the field of emotional vocal communication. Reply to comments on "Emotional voices in context: A neurobiological model of multimodal affective informati. *Physics of Life Reviews*, *9*(1), 9–12. http://doi.org/10.1016/j.plrev.2011.12.003

Butler, A., Oruc, I., Fox, C., & Barton, J. S. (2009). Factors contributing to the adaptation aftereffects of facial expression. *Brain*, *1191*, 116–126. http://doi.org/10.1016/j.brainres.2007.10.101.Factors

Byrd, M., Bowman, C., & Yamauchi, T. (2011). Cooing, crying, and babbling : A link between music and prelinguistic communication. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the 34th Annual Conference of the Cognitive Science Society* (pp. 1392–1397). Cognitive Science Society.

Cacioppo, J. T., Gardner, W. L., & Berntson, G. G. (1999). The affect system has parallel and integrative processing components: Form follows function. *Journal of Personality and Social Psychology*, *76*(5), 839–855. http://doi.org/10.1037/0022-3514.76.5.839

Caclin, A., Giard, M.-H., & McAdams, S. (2009). Perception of timbre dimensions: Psychophysics and electrophysiology in humans. *The Journal of the Acoustical Society of America*, *126*(4), 2236. http://doi.org/10.1121/1.3249185

Caclin, A., McAdams, S., Smith, B. K., & Winsberg, S. (2005). Acoustic correlates of timbre space dimensions: a confirmatory study using synthetic tones. *The Journal of the Acoustical Society of America*, *118*(1), 471–482. http://doi.org/10.1121/1.1929229

Carver, C. S., Sutton, S. K., & Scheier, M. F. (2000). Action, emotion, and personality: Emerging conceptual integration. *Personality and Social Psychology Bulletin*, *26*(6), 741–751. http://doi.org/10.1177/0146167200268008

Chartrand, J. P., Peretz, I., & Belin, P. (2008). Auditory recognition expertise and domain specificity. *Brain Research*, *1220*, 191–198. http://doi.org/10.1016/j.brainres.2008.01.014

Coutinho, E., & Deng, J. (2014). Transfer learning emotion manifestation Across music and speech, In *Neural Networks (IJCNN), 2014 International Joint Conference on* (pp. 3592-3598). IEEE.

Coutinho, E., & Dibben, N. (2012). Psychoacoustic cues to emotion in speech prosody and music. *Cognition & Emotion*, (July 2013), 1–27. http://doi.org/10.1080/02699931.2012.732559

Dalla Bella, S., Peretz, I., Rousseau, L., & Gosselin, N. (2001). A developmental study of the affective value of tempo and mode in music. *Cognition*, *80*(3), 1–10. http://doi.org/10.1016/S0010-0277(00)00136-0

Darwin, C. (1872). *The expression of the emotions in man and animals*.

Dowling, W. J. (1978). Scale and contour: Two components of a theory of memory for melodies. *Psychological Review*, *85*(4), 341–354. http://doi.org/10.1037/0033-295X.85.4.341

Drossos, K., Floros, A., & Kanellopoulos, N.-G. (2012). Affective acoustic ecology. *Proceedings of the 7th Audio Mostly Conference on A Conference on Interaction with Sound - AM '12*, (SEPTEMBER 2012), 109–116. http://doi.org/10.1145/2371456.2371474

Eder, a. B., Elliot, a. J., & Harmon-Jones, E. (2013). Approach and avoidance motivation: Issues and advances. *Emotion Review*, *5*(3), 227–229.

http://doi.org/10.1177/1754073913477990

Eerola, T., Ferrer, R., & Alluri, V. (2012). Timbre and affect dimensions:Evidence from affect and similarity ratings and acoustic correlates of isolated instrument sounds. *Music Perception*, *30*,(1), 49–70.

Eerola, T., Friberg, A., & Bresin, R. (2013). Emotional expression in music: Contribution, linearity, and additivity of primary musical cues. *Frontiers in Psychology*, *4*, 1–12. http://doi.org/10.3389/fpsyg.2013.00487

Eerola, T., & Vuoskoski, J. K. (2013). A review of music and emotion studies: Approaches, emotion, models, and stimuli. *Music Perception*, *30*(3), 307–340.

Ekman, P. (1992). An argument for basic emotions. *Cognition & Emotion*, *6(3-4)*, 169-200. http://doi.org/10.1080/02699939208411068

Ekman, P. (1992). Are there basic emotions? *Psychological Review*, *99*(3), 550–553. http://doi.org/10.1037/0033-295X.99.3.550

Ekman, P., & Friesen, W. (1986). A new pan-cultural facial expression of emotion.. *Motivation and Emotion*, *10*(2), 159–168.

Fedorenko, E., Patel, A., Casasanto, D., Winawer, J., & Gibson, E. (2009). Structural integration in language and music: evidence for a shared system. *Memory & Cognition*, *37*(1), 1–9. http://doi.org/10.3758/MC.37.1.1

Fernald, a. (1989). Intonation and communicative intent in mothers' speech to infants: Is the melody the message? *Child Development*, *60*(6), 1497–1510. http://doi.org/10.1111/1467-8624.ep9772504

Fox, C. J., & Barton, J. J. S. (2007). What is adapted in face adaptation? The neural representations of expression in the human visual system. *Brain Research*, *1127*(1), 80–89. http://doi.org/10.1016/j.brainres.2006.09.104

Fritz, T., Jentschke, S., Gosselin, N., Sammler, D., Peretz, I., Turner, R., … Koelsch, S. (2009). Universal Recognition of Three Basic Emotions in Music. *Current Biology*, *19*(7), 573–576. http://doi.org/10.1016/j.cub.2009.02.058

Frühholz, S., & Grandjean, D. (2013). Amygdala subregions differentially respond and rapidly adapt to threatening voices. *Cortex*, *49*(5), 1394–1403. http://doi.org/10.1016/j.cortex.2012.08.003

Gabrielsson, A., & Juslin, P. N. (1996). Emotional expression in music performance: Between the performers intention and the listeners' experience. *Psychology of*

Music, 24(1), 68-91.

Gabrielsson, A., & Lindstrom, E. (2010). The role of structure in the musical expression of emotions. In *Handbook of music and emotion: theory, research, applications* (pp. 367–400).

Gagnon, L., & Peretz, I. (2003). Mode and tempo relative contributions to "happy-sad" judgements in equitone melodies. *Cognition & Emotion*, *17*(1), 25–40. http://doi.org/10.1080/02699930302279

Grey, J. M., & Moorer. (1977). Multidimensional perceptual scaling of musical timbres. *The Journal of the Acoustical Society of America*, *61*(5), 1270–1277. http://doi.org/10.1121/1.381428

Grey, J. M., & Moorer, J. A. (1977). Perceptual evaluations of synthesized musical instrument tones. *The Journal of the Acoustical Society of America*, *62*(2), 454–462.

Grill-Spector, K., Kushnir, T., Edelman, S., Avidan, G., Itzchak, Y., & Malach, R. (1999). Differential processing of objects under various viewing conditions in the human lateral occipital complex. *Neuron*, *24*(1), 187–203. http://doi.org/10.1016/S0896-6273(00)80832-6

Hailstone, J. C., Omar, R., Henley, S. M. D., Frost, C., Kenward, M. G., & Warren, J. D. (2009). It's not what you play, it's how you play it: timbre affects perception of emotion in music. *Quarterly Journal of Experimental Psychology (2006)*, *62*(11), 2141–2155. http://doi.org/10.1080/17470210902765957

Hajda, J., Kendall, R., Carterette, E., & Harshberger, M. (1997). *Methodological issues in timbre research*. (J. Sloboda, Ed.). Hove, England: Psychology Press.

Harmon-Jones, E. (2003). Clarifying the emotive functions of asymmetrical frontal cortical activity. *Psychophysiology*, *40*(6), 838–848. http://doi.org/10.1111/1469-8986.00121

Harmon-Jones, E., Harmon-Jones, C., & Price, T. F. (2013). What is approach motivation? *Emotion Review*, *5*(3), 291–295. http://doi.org/10.1177/1754073913477509

Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: what is it, who has it, and how did it evolve? *Science (New York, N.Y.)*, *298*(5598), 1569–1579. http://doi.org/10.1126/science.298.5598.1569

Helmholtz, H. von. (1885). *On the sensations of tone as a physiological basis for the*

theory of music (2nd ed.). London: Longman.

Huber, D. E., & O'Reilly, R. C. (2003). Persistence and accomodation in short-term priming and other perceptual paradigms: Temporal segregation through synaptic depression. *Cognitive Science*, *27*(3), 403–430. http://doi.org/10.1016/S0364-0213(03)00012-0

Hunter, P. G., Schellenberg, E. G., & Schimmack, U. (2010). Feelings and perceptions of happiness and sadness induced by music: Similarities, differences, and mixed emotions. *Psychology of Aesthetics, Creativity, and the Arts*, *4*(1), 47–56. http://doi.org/10.1037/a0016873

Ilie, & Thompson. (2006). A comparison of acoustic cues in music and speech for three imensions of affect. *Music perception: An interdisciplinary journal*, *23*(4), 319–330. http://doi.org/10.1525/rep.2008.104.1.92.

Izard. (1977). *Human Emotions*. New York: Plenum Press.

Johnson-Laird, P., & Oatley, K. (1989). The language of emotions: An analysis of a semantic field. *Cognition and Emotion*, *3*(2), 81–123.

Juslin, P. (1997). Emotional communication in music performance: A functionalist perspective and some data. *Music Perception*, *14*(4), 383–418. Retrieved from http://psycnet.apa.org/?fa=main.doiLanding&uid=1997-05731-002\nhttp://www.jstor.org/stable/10.2307/40285731

Juslin, P. ., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: different channels, same code? *Psychological Bulletin*, *129*(5), 770–814. http://doi.org/10.1037/0033-2909.129.5.770

Juslin, P. N. (2000). Cue utilization in communication of emotion in music performance: relating performance to perception. *Journal of Experimental Psychology. Human Perception and Performance*, *26*(6), 1797–1813. http://doi.org/10.1037/0096-1523.26.6.1797

Juslin, P. N. (2013). What does music express? Basic emotions and beyond. *Frontiers in Psychology*, *4*, 1–14. http://doi.org/10.3389/fpsyg.2013.00596

Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: different channels, same code? *Psychological Bulletin*, *129*(5), 770–814. http://doi.org/10.1037/0033-2909.129.5.770

Juslin, P. N., & Scherer, K. R. (2005). *Vocal expression of affect*. (J. Harrigan, R.

Rosenthal, & K. R. Scherer, Eds.). New York: Oxford University Press.

Juslin, P. N., & Sloboda, J. A. (2001). *Music and emotion: Theory and research.*. Oxford University Press.

Kandel, E. R., & Siegelbaum, S. A. (2012). Cellular mechanisms of implicit memory storage and the biological basis of individuality. In E. Kandel, J. Schwartz, T. Jessell, S. Siegelbaum, & A. Hudspeth (Eds.), *Principles of Neural Science* (pp. 1461–1486). New York: McGraw HIll Education.

Kawahara, H., & Matsui, H. (2003). Auditory morphing based on an elastic perceptual distance metric in an interference-free time-frequency representation. *2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP '03).*, *1*(August). http://doi.org/10.1109/ICASSP.2003.1198766

Kent, R. D. (1997). *The speech sciences*. San Diego: Singular Publishing.

Klingbeil, M. (2005). Software for spectral analysis, editing, and synthesis. *Synthesis*, (1), 107–110. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.128.5913&rep=rep1&type=pdf

Klinge, C., Röder, B., & Büchel, C. (2010). Increased amygdala activation to emotional auditory stimuli in the blind. *Brain*, *133*(6), 1729–1736. http://doi.org/10.1093/brain/awq102

Koelsch, S. (2005). Neural substrates of processing syntax and semantics in music. *Music That Works: Contributions of Biology, Neurophysiology, Psychology, Sociology, Medicine and Musicology*, 143–153. http://doi.org/10.1007/978-3-211-75121-3_9

Koelsch, S. (2011). Toward a neural basis of music perception - a review and updated model. *Frontiers in Psychology*, *2*, 1–20. http://doi.org/10.3389/fpsyg.2011.00110

Kreifelts, B., Jacob, H., Brück, C., Erb, M., Ethofer, T., & Wildgruber, D. (2013). Non-verbal emotion communication training induces specific changes in brain function and structure. *Frontiers in Human Neuroscience*, *7*(October), 648. http://doi.org/10.3389/fnhum.2013.00648

Krumhansl, C. L. (1997). An exploratory study of musical emotions and psychophysiology. *Canadian Journal of Experimental Psychology*, *51*(4), 336–353.

http://doi.org/10.1037/1196-1961.51.4.336

Kwon, O., Chan, K., Hao, J., & Lee, T. (2003). Emotion recognition by speech signals. *Eighth European Conference on Speech Communication and Technology*, 125–128. http://ergo.ucsd.edu/~leelab/pdfs/ES030151.pdf

Lang, P. (1995). The emotion probe: Studies of motivation and attention. *American Psychologist*, *50*(5), 372–385.

Lartillot, O., Toiviainen, P., & Eerola, T. (2008). A matlab toolbox for music information retrieval. *Data Analysis, Machine Learning and Applications*, 261–268.

Laukka, P., Eerola, T., Thingujam, N. S., Yamasaki, T., & Beller, G. (2013). Universal and culture-specific factors in the recognition and performance of musical affect expressions. *Emotion*, *13*(3), 434–49. http://doi.org/10.1037/a0031388

Leman, M., Vermeulen, V., De Voogdt, L., Moelants, D., & Lesaffre, M. (2005). Prediction of musical affect using a combination of acoustic structural cues. *Journal of New Music Research*, *34*(1), 39–67. http://doi.org/10.1080/09298210500123978

Leopold, D. a, O'Toole, a J., Vetter, T., & Blanz, V. (2001). Prototype-referenced shape encoding revealed by high-level aftereffects. *Nature Neuroscience*, *4*(1), 89–94. http://doi.org/10.1038/82947

Leung, J. (2014). Timbre Recognition.[Powerpoint slides] Retrieved from http://www.music.mcgill.ca/~jason/mumt621/621_Presentation4.pdf

Levy, D. a, Granot, R., & Bentin, S. (2001). Processing specificity for human voice stimuli: electrophysiological evidence. *Neuroreport*, *12*(12), 2653–2657. http://doi.org/10.1097/00001756-200108280-00013

Liaw, a, & Wiener, M. (2002). Classification and regression by randomForest. *R News*, *2*(December), 18–22. http://doi.org/10.1177/154405910408300516

Logan, K. J. (2001). The effect of syntactic complexity upon the speech fluency of adolescents and adults who stutter. *Journal of Fluency Disorders*, *26*(2), 85–106. http://doi.org/10.1016/S0094-730X(01)00093-6

Loughran, R., Walker, J., O'Neill, M., & O'Farrell, M. (2004). The use of mel-frequency cepstral coefficients in musical instrument identification. *Proceedings of the International Computer Music Conference*, 42–43.

MacLin, O. H., Nelson, H. A., & Webster, M. a. (1996). Figural after-effects in the

perception of faces. *Investigative Ophthalmology and Visual Science*, *37*(3), 647–653. http://doi.org/10.3758/BF03212974

McAdams, S., & Bigand, E. (1993). Introduction to auditory cognition. *Thinking in Sound*, 1–8. http://interactive.colum.edu/mtd2/emily/readings/Intro_to_Auditory_Cognition.pdf

McAdams, S., & Cunible, J. C. (1992). Perception of timbral analogies. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *336*(1278), 383–389. http://doi.org/10.1098/rstb.1992.0072

McAdams, S., Winsberg, S., Donnadieu, S., De Soete, G., & Krimphoff, J. (1995). Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes. *Psychological Research*, *58*(3), 177–192. http://doi.org/10.1007/BF00419633

McDermott, J. H., & Oxenham, A. J. (2008). Music perception, pitch, and the auditory system. *Current Opinion in Neurobiology*, *18*(4), 452–463. http://doi.org/10.1016/j.conb.2008.09.005.Music

Mehrabian, a, & Wiener, M. (1967). Decoding of inconsistent communications. *Journal of Personality and Social Psychology*, *6*(1), 109–114. http://doi.org/10.1037/h0024532

Mehrabian, A., & Ferris, S. (1967). Inference of attitudes from nonverbal communication in two channels. *Journal of Consulting Psychology*, *31*(3), 248.

Meyer, L. B. (1956). *Emotion and Meaning in Music*. Chicago: University of Chicago Press.

Miell, D., Macdonald, R., Hargreaves, D., & Cross, I. (2004). Music and meaning, ambiguity and evolution. *Musical Communication*, 27–44. http://doi.org/10.1093/acprof:oso/9780198529361.003.0002

Neiberg, D., Elenius, K., & Laskowski, K. (2006). Emotion recognition in spontaneous speech using GMMs. *Proceedings ICSLP-2006, Pittsburgh*, 809 – 812. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.66.9789&rep=rep1&type=pdf

Oller, D. K. (2000). *The emergence of the speech capacity*. Mahwah, N.J.: Lawrence Erlbaum.

Padova, A., Bianchini, L., Lupone, M., & Belardinelli, M. O. (2003). Influence of

specific spectral variations of musical timbre on emotions in the listeners. *Proceedings of the 5th Triennial ESCOM Conference*, (September), 227–230.

Patel, A. (2009). *Music and the brain: Three links to language*. *The Oxford handbook of music psychology*.

Patel, A. D. (2003). Language, music, syntax and the brain. *Nature Neuroscience*, *6*(7), 674–681. http://doi.org/10.1038/nn1082

Patel, A. D. (2007). *Music, Language, and the Brain* (Vol. 5). Oxford University Press. https://books.google.com/books?hl=en&lr=&id=BDS2OSt1G-MC&pgis=1

Patil, K., Pressnitzer, D., Shamma, S., & Elhilali, M. (2012). Music in our ears: The biological bases of musical timbre perception. *PLoS Computational Biology*, *8*(11). http://doi.org/10.1371/journal.pcbi.1002759

Pell, M., Monetta, L., Paulmann, S., & Kotz, S. A. (2009). Recognizing emotions in a foreign language. *Journal of Nonverbal Behavior*, *33*(2), 107–120.

Peretz, I., Gagnon, L., & Bouchard, B. (1998). Music and emotion: perceptual determinants, immediacy, and isolation after brain damage. *Cognition*, *68*(2), 111–141. http://doi.org/10.1016/S0010-0277(98)00043-2

Peretz, I., Radeau, M., & Arguin, M. (2004). Two-way interactions between music and language: evidence from priming recognition of tune and lyrics in familiar songs. *Memory & Cognition*, *32*(1), 142–152. http://doi.org/10.3758/BF03195827

Plomp, R., & Levelt, W. J. (1965). Tonal consonance and critical bandwidth. *The Journal of the Acoustical Society of America*, *38*(4), 548–560. http://doi.org/10.1121/1.1909741

Quinto, L., Thompson, W. F., & Taylor, a. (2013). The contributions of compositional structure and performance expression to the communication of emotion in music. *Psychology of Music*, *42*(4), 503–524. http://doi.org/10.1177/0305735613482023

Rhodes, G., Brennan, S., & Carey, S. (1987). Identification and ratings of caricatures: implications for mental representations of faces. *Cognitive Psychology*, *19*(4), 473–497. http://doi.org/10.1016/0010-0285(87)90016-8

Risset, J. C., & Wessel, D. L. (1982). Exploration of timbre by analysis and synthesis. *The Psychology of Music*, *2*, 151.

Roesch, E. B., Korsten, N., Fragopanagos, N. F., Taylor, J. G., Grandjean, D., & Sander, D. (2011). Biological computational constraints to the psychological modelling of

emotion. In *Emotion-Oriented Systems* (pp. 47-62). Springer Berlin Heidelberg. http://doi.org/10.1007/978-3-642-15184-2

Rousseau, J.-J., & von Herder, J. G. (1986). *On the origin of language*. University of Chicago Press.

Russell, J. a, & Carroll, J. M. (1999). On the bipolarity of positive and negative affect. *Psychological Bulletin*, *125*(1), 3–30. http://doi.org/10.1037/0033-2909.125.1.3

Schachner, A., & Hannon, E. E. (2011). Infant-directed speech drives social preferences in 5-month-old infants. *Developmental Psychology*, *47*(1), 19–25. http://doi.org/10.1037/a0020740

Scherer, K. R. (1995). Expression of emotion in voice and music. *Journal of Voice : Official Journal of the Voice Foundation*, *9*(3), 235–248. http://doi.org/10.1016/S0892-1997(05)80231-0

Scherer, K. R., & Oshinsky, J. S. (1977). Cue utilization in emotional attribution from auditory atimuli. *Motivation and Emotion*, *1*(4), 331–346.

Schubert, E. (2004). Modeling perceived emotion with continuous musical features. *Music Perception*, *21*(4), 561–585. http://doi.org/10.1525/mp.2004.21.4.561

Schweinberger, S. R., Casper, C., Hauthal, N., Kaufmann, J. M., Kawahara, H., Kloth, N., … Zäske, R. (2008). Auditory adaptation in voice perception. *Current Biology*, *18*(9), 684–688. http://doi.org/10.1016/j.cub.2008.04.015

Sloboda, J., & O'Neill, S. (2001). Emotions in everyday listening to music. In Juslin, P., & Sloboda, J. (Eds.) *Music and emotion: Theory and research* (415-429). New York, NY: Oxford University Press. http://psycnet.apa.org/psycinfo/2001-05534-012

Springer, U. S., Rosas, A., McGetrick, J., & Bowers, D. (2007). Differences in startle reactivity during the perception of angry and fearful faces. *Emotion*, *7*(3), 516–525. http://doi.org/10.1037/1528-3542.7.3.516

Strauss, M. M., Makris, N., Aharon, I., Vangel, M. G., Goodman, J., Kennedy, D. N., … Breiter, H. C. (2005). fMRI of sensitization to angry faces. *NeuroImage*, *26*(2), 389–413. http://doi.org/10.1016/j.neuroimage.2005.01.053

Terwogt, M., & Van Grinsven, F. (1991). Musical expression of moodstates. *Psychology of Music*, *19*, 99–109. http://doi.org/0803973233

Thompson, W. F., Schellenberg, E. G., & Husain, G. (2004). Decoding speech prosody:

do music lessons help? *Emotion*, *4*(1), 46–64. http://doi.org/10.1037/1528-3542.4.1.46

Trehub, S. E., Becker, J., & Morley, I. (2015). Cross-cultural perspectives on music and musicality. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *370*(1664), 20140096..

Wallbott, H. G., & Scherer, K. R. (1986). Cues and channels in emotion recognition. *Journal of Personality and Social Psychology*, *51*(4), 690–699. http://doi.org/10.1037/0022-3514.51.4.690

Watson, D., Wiese, D., Vaidya, J., & Tellegen, A. (1999). The two general activation systems of affect: Structural findings, evolutionary considerations, and psychobiological evidence. *Journal of Personality and Social Psychology*, *76*(5), 820–838. http://doi.org/10.1037/0022-3514.76.5.820

Webster, M. A., Kaping, D., Mizokami, Y., & Duhamel, P. (2004). Adaptation to natural facial categories. *Nature*, *428*, 557–561. http://doi.org/10.1038/nature02361.1.

Weninger, F., Eyben, F., Schuller, B. W., Mortillaro, M., & Scherer, K. R. (2013). On the acoustics of emotion in audio: What speech, music, and sound have in common. *Frontiers in Psychology*, *4*(MAY), 1–12. http://doi.org/10.3389/fpsyg.2013.00292

Wilkowski, B. M., & Meier, B. P. (2010). Bring it on: angry facial expressions potentiate approach-motivated motor behavior. *Journal of Personality and Social Psychology*, *98*(2), 201–210. http://doi.org/10.1037/a0017992

Zeng, Z., Pantic, M., Roisman, G. I., & Huang, T. S. (2009). A survey of affect recognition methods: Audio, visual, and spontaneous expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *31*(1), 39–58. http://doi.org/10.1109/TPAMI.2008.52