PERCEIVING EMOTION IN SOUNDS: DOES TIMBRE PLAY A ROLE?

A Thesis

by

CASADY DIANE BOWMAN

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

MASTER OF SCIENCE

December 2011

Major Subject: Psychology

Perceiving Emotion in Sounds: Does Timbre Play a Role?

PERCEIVING EMOTION IN SOUNDS: DOES TIMBRE PLAY A ROLE?

A Thesis

by

CASADY DIANE BOWMAN

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

MASTER OF SCIENCE

Approved by:

Chair of Committee,	Takashi Yamauchi
Committee Members,	Jyostna Vaid
	Jayson Beaster-Jones
Head of Department,	Ludy Benjamin

December 2011

Major Subject: Psychology

ABSTRACT

Perceiving Emotion in Sounds: Does Timbre Play a Role? (December 2011)

Casady Diane Bowman, B.S., Oklahoma State University

Chair of Advisory Committee: Dr. Takashi Yamauchi

Acoustic features of sound such as pitch, loudness, perceived duration and timbre have been shown to be related to emotion in regard to sound, demonstrating that research involving the important connection between the perceived emotions and their timbres is lacking. This study investigates the relationship between acoustic features of sound and emotion with regard to timbre. In two experiments, we investigated whether particular acoustic components of sound can predict timbre and particular categories of emotion, and how these attributes are related. Two behavioral experiments related perceived emotion ratings with synthetically created sounds and International Affective Digitized Sounds (IADS). Also, two timbre experiments found a connection between acoustic components of synthetically created sounds, and IADS. Regression analyses uncovered some relationships between emotion, timbre, and acoustic features of sound. Results indicate that emotion is perceived differently for synthetic instrumental sounds and IADS. Mel-frequency cepstral coefficients were a strong predictor of perceived emotion of instrumental sounds; however, this was not the case for the IADS. This difference lends itself to the idea that there is a strong relationship between emotion and timbre for instrumental sounds, perhaps in part because of their relationship to speech and the way these different sounds are processed.

.

ACKNOWLEDGEMENTS

I would like to thank my committee chair, Dr. Yamauchi, and committee members, Dr. Vaid, and Dr. Beaster-Jones, for their support throughout the course of this research.

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

1. INTRODUCTION

Music and language are the most cognitively complex and emotionally expressive sounds invented by humans. They are both generative; that is, complexity is built up by rules and hierarchical organization that result in sentences or songs. So what is it that links these two modes of communication? Much is known and studied about the syntactic relations between music and language, but is there more we can say based on their sound relations, emotion, or how we use them? The study of music used as a form of emotion may help to disentangle the mysteries of its use in social communication, as well as the functional dissimilarities and similarities. Research distinguishing between music and language, and finding a link between timbre and emotion, can help to further identify the role of the processes for music and language in the brain.

This study focuses on the relationship between timbre and emotion. There is much research regarding timbre (Koelsch, 2005), but few studies have explored the link between timbre and emotion, (see Caclin et al., 2006, and Hailstone et al., 2009 for exceptions) to any degree of specificity. The main question addressed here is, do particular acoustic components of sound predict particular categories of emotion (e.g., happiness, sadness, anger, fear or disgust; see Ekman, 1992), as well as timbre? Perceiving timbre is presumed to rely upon the capacity to perceive and process differences between sounds, such as the difference between musical instruments or voices.

This thesis follows the style of the Journal of the Acoustical Society of America.

This ability to distinguish between sounds is essential to everyday human functioning, and is a fundamental task of the auditory system (McAdams & Cunible, 1992; Godyke et al., 2003). But how is this capacity linked to our ability to perceive emotions? If music and speech share some fundamental characteristics, then the ability to perceive timbre should be also related to the ability to perceive speech sounds. By investigating the relationship between timbre and emotion, this research aims to shed light on the basic acoustic features that define it.

The outline of this thesis is as follows. Related research analyzing timbre, emotion and the link between timbre and emotion is reviewed in sections 1.1 – 1.3. Section 1.4 gives an overview of experiments, and 1.5 details computational sound analyses for timbre extractions. In section 1.6 correlations of acoustic components are discusses, followed by 2.0 which details predictions of the data. Section 3.0 includes two experiments that demonstrate, and explain the similarities between timbre and emotion in terms of acoustic features. In section 3.3 and 3.4 principal component analysis is reviewed. Section 3.5 includes a preliminary data analysis, of Experiments 1a, and 1b as well as their results and discussion in section 3.8. Section 4.0 comprises Experiments 2a and 2b as well as their results, and discussion. Finally, section 5.0 consists of a general discussion section. Overall, this research aims to investigate and further explicate the relationship between timbre, sound, and emotion.

1.1. Timbre

Sounds are perceived and characterized by a number of attributes such as pitch, loudness, perceived 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 (Hailstone et al., 2009; Bregman, Liao & Levitan, 1990; McAdams & Cunible, 1992). When identifying a musical instrument, to tell the difference between a flute and guitar playing the same note, pitch, and duration, one uses timbre. This quality of timbre allows a listener to identify individual instruments of an orchestra, and involves dynamic features of the sound, especially onset characteristics (Grey & Moorer, 1977, and Risset & Wessel, 1982).

The classic definition of timbre holds that different timbres result from the sound of different amplitudes (of harmonic components) of a complex tone in a steady state" (Helmholtz, 1885). These definitions illustrate the relationship between sound and timbre in that it is a feature of sound, but they do not adequately inform us regarding acoustic components that create different timbres and how these components are shared for the perception of emotions of sounds. Timbre is complex and is made up of several acoustic components; it is multidimensional (Caclin et al., 2005). The multidimensionality of timbre makes it difficult to study or measure on a single continuum such as low to high. Contrary to pitch, which relies on the tone's fundamental frequency and loudness, timbre relies on several parameters, or acoustic dimensions of the sound. The main goal of most timbre studies has been to uncover the number and nature of these dimensions of timbre. A method most often used is that of multidimensional scaling (MDS) of dissimilarity ratings (McAdams & Bigand, 1993; Hajda et al., 1997). The advantage of MDS is that it does not make any assumptions about the acoustic dimensions of a sound. Studies using MDS typically have listeners rate the dissimilarity between two stimuli, to result in a dissimilarity matrix which undergoes MDS 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 et al., 1995) in order to better understand how the perception of timbre, and emotion are related.

In the research of Padova et al., (2003), an often misled notion is discussed that sounds with identical spectra, or sound distribution, have identical timbres. Berger (1964) notes that the timbre of a piano tone is perceived as completely different when it is played backward even though the original and the reversed sound have the same spectra (Berger 1964). Another point of interest is that even major changes of the spectrum of a tone, do not prevent a listener from recognizing a musical instrument. Padova et al. (2003) argue that musical timbre does not depend upon one single physical dimension. Other researchers (see Caclin et al., 2005; Hailstone et al., 2009) have shown that other features such as amplitude, phase, attack time, and decay in a tone, all work simultaneously to influence the perception of timbre.

Some of the most studied populations are those of musicians in regard to their music processing abilities. Musicians are able to outperform non-musicians when processing an instrument's timbre. Research such as that of McAdams et al., (1995) has evaluated the perceptual structure of musical timbre in musicians, amateur musicians and non-musicians. Using a three-dimensional model, McAdams et al. (1995) was able to identify the attack time, the spectral centroid, and the spectral flux to be the acoustic correlates to discriminate timbres in a dissimilarity-rating task (Chartrand & Belin, 2008).

Recent studies using a multidimensional scaling technique have identified two-dimensional and three dimensional structures of timbre (Rasch & Plomp 1982; Wedin & Goude 1972; Wessel & Grey 1978; Grey & Moorer, 1977; Miller & Carterette, 1975; Krumhansl, 1989; Plomp, 1970; McAdams & Cunible, 1992). The timbre space resulting from the studies of Miller and Carterette (1975) discovered a three dimensional model where two of the dimensions were related to the harmonic structure, and the third was related to the amplitude envelope of a sound; similar results were achieved by Samson, Zatorre & Ramsay (1997). According to the research and studies of Grey & Moorer (1977), three dimensions exist for describing timbre, that of spectral energy distribution, presence of synchronicity also termed spectral fluctuation and the presence of low-amplitude, high-frequency energy in the attack of a sound.

Furthermore, several studies highlight the role of the distribution of spectral energy in dissimilarity and similarity judgments (Plomp 1970; Samson, Zatorre & Ramsay, 1997; Wedin and Goude 1972; Grey & Moorer, 1977; Krumhansl 1989). Music producers and researchers alike are now able to produce and create many kinds of complex sounds by controlling for specific acoustical properties (Padova et al., 2003). While this applies to music excerpts, and pieces, the timbral variations within a single instrument that are used to transmit emotional expressions are different and are likely smaller than those that are present between instruments (Godyke et al., 2003).

In summary, research in the field of timbre shows that there are several acoustic features with which timbre can be characterized such as attack time, and spectral flux; however, these do not allow for the full range of emotion that is said to describe timbre.

1.2. Emotion

Emotions are social. To understand the relationship between emotions and the social world, it is necessary to include a social psychological approach. To say that emotions are social is to say that emotions are deeply entrenched in our social world. For example, we experience jealousy in relationships, appreciation for help from others, and anger at others actions. It is also appropriate to look at social roles people play in interactions – these can specify what emotions and moods are to be displayed in a given situation. It has been argued that the communication of emotions serves as the groundwork of the social order in animals and humans. However, this same type of communication is also significant within performing arts such as music.

The scope of this present research will make use of “basic” emotions. Ekman (1992) states that the meaning of “basic” emotions illuminates the viewpoint that emotions have evolved for their adaptive value in dealing with fundamental life tasks. These fundamental life tasks as described by Johnson-Laird & Oatley (1992) are universal human predicaments, such as achievements, losses, frustrations, etc. These basic emotions, and fundamental life tasks, are adaptive in that they lead us, in the course of evolution, to create better solutions than those used previously in attaining relevant goals. Emotions deal with recurrent adaptive situations, (Tooby & Cosmides, 1990), these adaptive situations emphasize what distinguishes emotions, our appraisal of a current event is influenced by our ancestral past.

Emotions represent reactions to an event of significance; they produce changes in an organism. One important feature of emotion is that it produces specific action readiness (or reactions) while providing a latency period to allow for adaptation of behavioral reactions to a situation (Scherer, 1995). This latency period is used so that the organism can predict the reaction of others to an action as the result of a particular emotional state. As in the classic work of emotion in humans and animals by Darwin (1872), it has been shown that emotional expressions provide an essential function of communicating action and reaction to the social environment (Scherer, 1995). Emotion as well as expression are phylogenetically continous and are found in many species, especially in species where social life is based on complex interactions between individuals. Many expressive modalities are important to emotion communication such as body posture, facial features, and vocalization (Scherer, 1995). Communication of emotions is crucial to social relationships and survival (Ekman, 1992). The two most effective resources for emotional communication are both vocal expression and music (Gabrielsson & Juslin, 1996).

It is clear that emotions in music are important; yet there are issues that remain difficult to resolve such as whether music can convey specific emotions, or if music really does evoke emotion in listeners. Facial recognition research by Ekman (1992) showed that the basic emotions, happy, sad, anger, and fear, are universal and crosscultural, as well as important for social communication. Such emotions are also prevalent within music and sound used for communication.

To summarize, the past research on the relationship between music and emotions has well covered the association between social cognition, universality, and physiological arousal explained within emotion; however, very little research has covered the link between emotion and sound in terms of acoustic components (see Caclin et al. 2006; and Hailstone et al., 2009 for a few exceptions). With the exception of work by Bradley and Lang (1999), using the International Affective Digitized Sounds (IADS), most studies have not examined the connection between sound and emotion in terms of important acoustic components that work to explain emotion in sound.

1.3. Linking timbre and emotion

Distinct sounds in both language and music are used to express emotion, but what acoustic features of sound relate to emotion? Previous research has shown that emotion in music and sound is influenced by structure such as melodic contour, vibrato, tempo, rhythm, mode, consonance, dissonance and timbre (Gabrielsson & Juslin, 1996). Listeners are able to readily interpret emotional meaning of music by attending to specific properties of the music (Hevner, 1935; Balkwill et al., 2004). As an example, joy in music is often associated with fast tempo, a major mode, wide pitch range, and high loudness (Gabrielsson & Juslin, 1996). These properties of music, such as tempo and loudness could provide evidence to support universal cues to emotion in music. Such acoustic cues are used, either unconsciously or consciously by performers and composers (Balkwill et al., 2004) as well as culturally specified conventions to determine and express emotion in music.

Other studies have also proposed that emotion in music is related to pitch, tempo, loudness and timbre of speech (Ladd et al., 1985; Johnstone & Scherer, 2000). Though specifics of studies are different, both domains of research suggest timbre as one important factor for experiencing emotion in speech, music, and sound.

The emotions happy and anger are similar in terms of acoustic cues relating to rate, intensity and pitch patterns, yet differ in regard to timbre (Patel, 2009). Hailstone et al. (2009) found that instrument identity, or timbre, influences perception of emotion in music. Other early studies such as those by Hajda et al., (1997) demonstrated the use of timbre's temporal and spectral components in instrument recognition. This was done using recorded and transformed versions of sounds. Results showed that both spectral and temporal characteristics were important for an instrument recognition task. This demonstrates the importance of giving further attention to studying timbre as a major contributor to emotion in music.

A major limitation of past research on timbre and music is that there is little focus on the relationship between the perceptual components of timbre and perceived emotion. Adding to the drawback are the differing claims that have been made with reference to emotion in music. For example, it has been stated that emotions are spontaneous responses, or that emotions are consistent between subjects, or that music does not induce basic emotions (Koelsch, 2005; Scherer, 2003). There is lacking in current research an important aspect connecting perceived emotion influenced by timbre, and sound identity (Hailstone et al., 2009). This gap in the literature reflects the idea that musical emotions are not like other emotions (Krumhansl, 1997). Differences between these emotions are evident in the antecedents and consequences of emotions. Antecedents are environmentally determined conditions that have perceived or real implications for an individual's welfare; these are commonly trailed by withdrawal or aggression, for example (Krumhansl, 1997). In order to physically prepare an individual to perform such actions, emotions are essential. Music however does not have such an overt effect on an individual's welfare; it is not often followed by a goal-directed action.

Here, the strategy is to investigate how acoustic components relate to timbre and emotion, both with synthetically created sounds, as well as with the International Affective Digitized sounds (Bradley & Lang, 2007). In conducting this study it was important to control for factors such as pitch, familiarity, and structural cues that could affect perception of emotion. Novel stimuli were created from ten instruments for Experiments 1a and 1b, Experiments 2a and 2b utilized the International Affective Digitized Sounds (IADS) (Bradley & Lang, 1999). Two, two-part experiments were conducted; for Experiment 1a and 1b sounds had synthetically modified timbres, these sounds were designed to include timbral cues to particular basic emotions. The basic emotions, happy, sad, anger and fear were chosen over other emotions because they support work on emotion perception from facial expressions by Ekman (1992), which shows that these emotions are universally recognized by normal human participants, and they are well represented in music.

Once ratings were obtained for Experiments 1a and 1b, the goal of analysis was to uncover the relationship between emotion, and timbre in the synthetically created sounds. This was done using principal components analysis to reduce the dimensions of the original data sets (both for acoustic components as well as emotion and timbre). A regression analysis was then applied to identify the acoustic components that would predict timbre and emotion ratings in sound. This same method of analysis was repeated for Experiments 2a and 2b using the IADS, which were more environmentally based sounds.

1.4. Overview of the experiments

The main question this research asks is if particular acoustic qualities of sound can explain, or predict particular categories of emotion and timbre. This research endeavors to find how these attributes of sound, timbre, and emotion are related. Two behavioral experiments were conducted: an instrument judgment experiment and an emotion judgment experiment as well as an analysis of previously collected data using the International Affective Digitized Sounds (IADS), (Bradley & Lang, 2007). Computational sound analyses were run on all sound stimuli. In the behavioral experiments, participants rated the extent to which instrumental, or IAD sounds conveyed particular emotions, timbre, or categories using a 1-7 scale.

To identify the acoustic properties that were able to predict instrument judgments and emotion judgments, eight components of timbre (i.e. *attack time, attack slope, zero-cross, roll off, brightness, mel-frequency cepstral coefficients, roughness, and irregularity*) were extracted from a total of 179 stimuli as well as 106 IADS stimuli. By applying principal component analysis (PCA), and stepwise multiple regression analyses, we compared which acoustic features of timbre could predict the behavioral performance obtained from the instrument judgment, the emotion judgment experiment and the IADS emotion data. Principal components analysis (PCA) was used on emotion, instrument, and International affective digitized sounds data to reduce the dimensionality.

To analyze the sounds and rating data, several different independent and dependent variables for the regression analyses were investigated. The first regression analysis uses the independent variable of predictors (acoustic features) as well as the dependent variable of emotion ratings for synthesized sounds (Experiment 1a). The next regression analysis uses the predictor variables (independent variables) and timbre ratings of the synthesized sounds (Experiment 1b). The same analyses were used with the independent variables for Experiments 2a and 2b, involving the International Affective digitized category ratings, as well as emotion ratings, respectively.

This study shows that there is a visible overlap as well as disparity in the acoustic components that explain timbre and emotion; this is most noted for the components mel-frequency cepstral coefficient for synthetically created instrumental sounds. Mfcc's have been especially important in the field of speech recognition; they are a set of perceptually motivated features that offer a condensed representation of the spectral envelope, such that most of the signal energy is concentrated in the first coefficients (Tzanetakis, 2002). For both the timbre and emotion judgments, these speech-related audio-features play a central role. However, this is not the case for the IADS.

1.5. Timbre extraction

In what follows, acoustic features of timbre are described in detail, as well as the computational procedure of extracting these features. The purpose of using these acoustic features is to act as predictors in regression analyses that can explain perceived emotion and timbre perception. A total of 179 sound stimuli were analyzed.

Eight acoustic properties of timbre: attack time, attack slope, zero-cross, roll off, brightness, mel-frequency cepstral coefficients, roughness, and irregularity were extracted from a total of 179 stimuli sounds using MIRToolbox in Matlab (Lartillot, Toiviainen, & Eerola, 2008). These acoustic properties are known to 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). The acoustic features were extracted from synthesized sounds rated in Experiments 1a for timbre, and 1b for emotion, as well as the IADS rated in Experiments 2a for category, and 2b for emotion.

Attack time is the time in seconds it takes for a sound to travel from amplitude of zero, to the maximum amplitude of a given sound signal, or more simply the temporal duration. Some features of timbre such as attack time contribute to the perception of emotion in music (Gabrielsson & Juslin, 1996; Juslin, 2000; Loughran et al., 2001); which suggests that features of timbre can at least in part determine the emotional content of music (Hailstone et al., 2009).

Attack time is computed using the equation of a line, $y = mx + b$, it is part of a sounds 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, Figure 1 gives a demonstration of attack time. The horizontal segments below the x-axis indicate the time it takes in seconds to achieve the maximum peak of each frame for which the attack time was calculated.

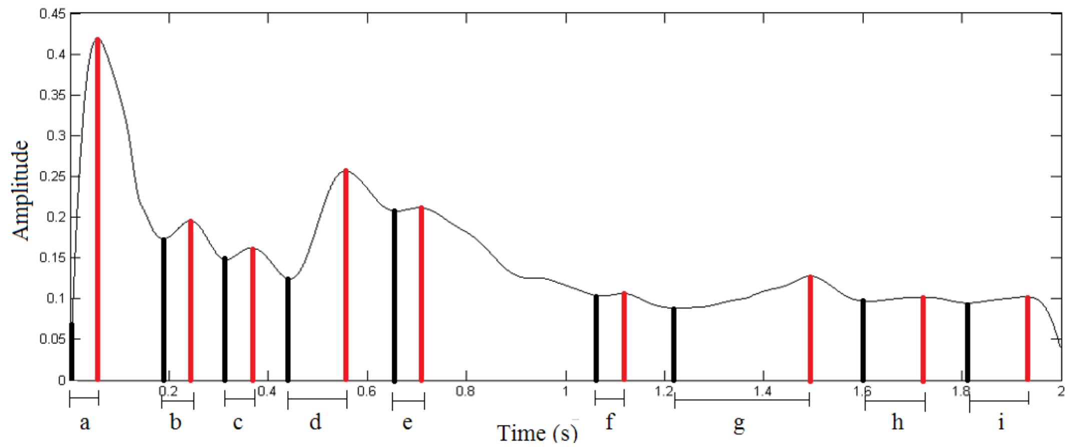


Figure 1. Attack time of a waveform audio file. This figure gives an example of the acoustic component attack time, for a waveform audio file (wav). Sections *a* through *i* in the figure indicates separate attack times; this is the time in seconds from the vertical black line, to the peak of the sound indicated by the vertical red line.

Attack slope is the attack phase of the amplitude envelope of a sound, also interpreted as the average slope leading to the attack time. This can also be calculated using the equation of a line y = mx +b, where *m* is the slope of the line and *b* is the point where the line crosses the vertical axis (*t*=0), see Figure 2. The red line in Figure 2 indicates the slope of the attack.

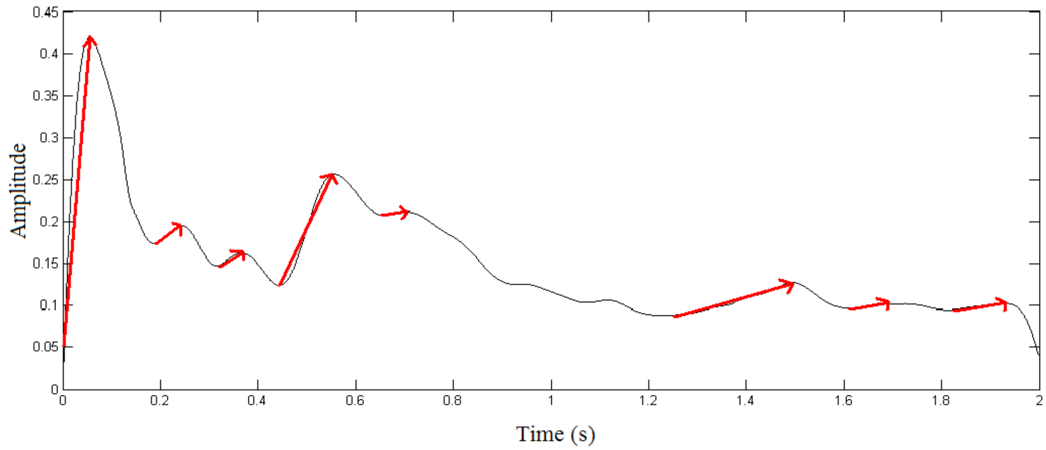


Figure 2. Attack slope of a waveform audio file. This figure gives an illustration of the acoustic component attack slope. The red arrow indicates the duration (attack time) for which the attack slope is calculated.

Zero-cross is the number of times a sound signal crosses the x-axis, this accounts for noisiness in a signal and is calculated using the following equation where sign is 1 for positive arguments and 0 for negative arguments. X[n] is the time domain signal for frame t.

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

Roll off is the amount of high frequencies in a signal which is specified by a cut-off point. The roll-off frequency is defined as the frequency where response is reduced by -3 dB. This is calculated using the following equation where Mt is the magnitude of the Fourier transform at frame *t* and frequency bin *n*. Rt is the cutoff frequency, see Figure 3.

$$\sum_{n=1}^{R_t} M_t[n] = 0.85 * \sum_{n=1}^{R_t} M_t[n]$$

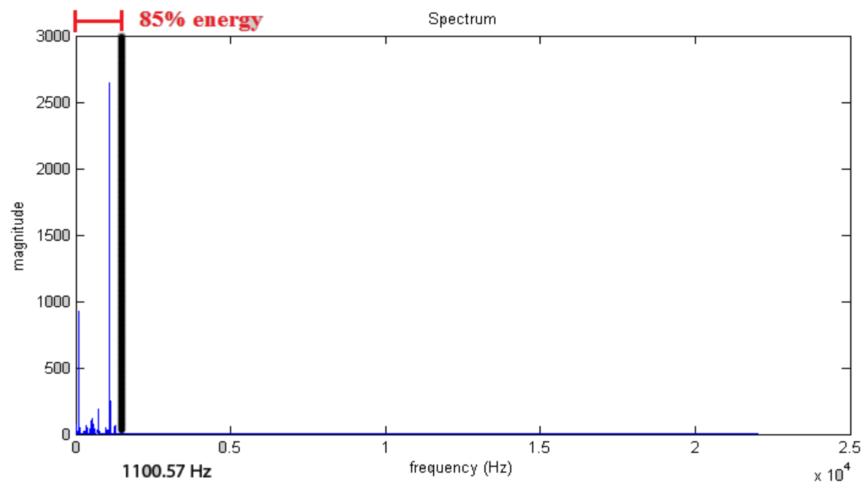


Figure 3. Roll off of a waveform audio file. This figure shows the acoustic component roll off, the red segment indicates the cutoff point of 85% for the amount of high frequencies in the signal.

Brightness is the amount of energy above a specified frequency, typically set at 1500 Hz – this is related to spectral centroid. The term "brightness" is also used in discussions of sound timbres, in a rough analogy with visual brightness. Timbre researchers consider brightness to be one of the strongest perceptual distinctions between sounds. Acoustically it is an indication of the amount of high-frequency content in a sound, and uses a measure such as the spectral centroid, see Figure 4.

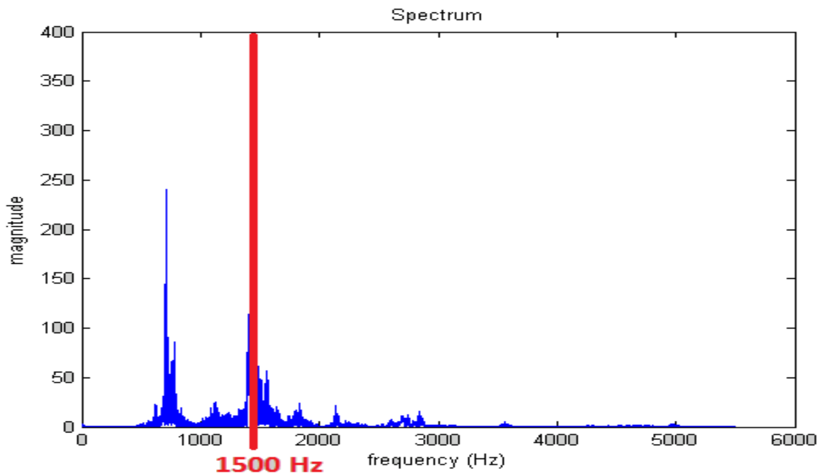


Figure 4. Brightness of a waveform audio file. This figure shows the acoustic component brightness. To the right of the red line is the amount of energy above 1500 Hz, or the brightness of the sound.

Roughness is sensory dissonance, the perceived harshness of a sound; this is the opposite of consonance (harmony) within music or even a 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.

Formally, roughness is calculated using the following equation where a_j and a_k are the amplitudes of the components, and g (f_{cb}) is a 'standard curve.' This was first proposed by Plomp & Levelt (1965).

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

Following extraction of the value for roughness from the sound stimuli, principal components analysis was used to reduce the dimensions of the roughness data, principal components analysis is explained in detail, in section 2.2.

Mel-frequency Cepstral Coefficients (mfcc) 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 the dominant features used in speech recognition as well as some music modeling (Logan, 2001). Frequencies in the Mel scale are equally spaced, and approximate the human auditory system more closely than a linearly spaced frequency bands used in a normal cepstrum. Due to large data output, prior to analyses mfcc data were reduced using principal components analyses to create a workable set of data. A cutoff criterion of 80% was used to represent the variability in the original mfcc data. Figure 5 shows the numerical Mel-frequency cepstral coefficient rank values for the 13 mfcc components returned. Thirteen components are returned due to the concentration of the signal information in only a few low-frequency components.

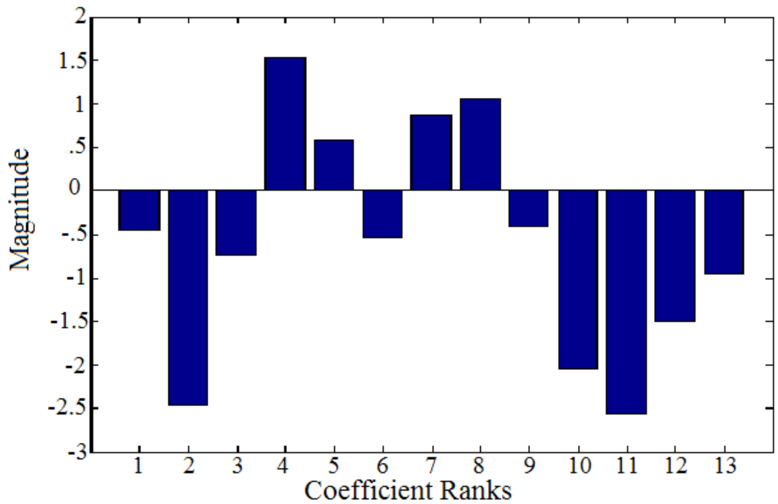


Figure 5. Mel-frequency cepstral coefficients (mfcc) of a waveform audio file. This figure shows the acoustic component mfcc. Each bar represents the numerical (rank coefficient) value computed for the thirteen components returned.

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

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

1.6. Correlation of acoustic components

This section reviews correlations found between acoustic components used as predictors in the regression analyses. To assure that the regressions of the principal components are run correctly, it is important to test for multicollinearity. In regression analysis, this is when two or more predictor variables in a multiple regression model are highly correlated. This can cause problems for the data in that calculations of individual predictors might not predict the data as well, while the predictive power (reliability) of the regression model as a whole is not reduced.

Table 1 shows the entire matrix of correlations (Pearson's *r*) among the fourteen predictors. This can give us an idea of how the emotion and timbre data will interact in terms of the predictors, or acoustic features.

Significantly correlated predictor variables include attack slope, with roughness, and zero cross; brightness with mfcc 2, mfcc 3, mfcc 6, roughness, zero cross, and roll off; irregularity with mfcc 2, mfcc 7, roughness and zero cross. These significant correlations indicate that the predictors used may not *individually* adequately predict timbre, or emotion. This means that none of the correlated predictors may contribute significantly to the model after the other one is included; however, altogether they contribute a lot. If the correlated variables are removed from the model, the fit of the model to the data will decrease. Simply put, it is possible that the overall model will fit the data, but that none of the correlated variables will have a significant contribution when added to the model.

Table 1

Pearson correlation of predictor variables

	Attack time	Attack slope	Bright-ness	Irregularity	MFCC 1	MFCC 2	MFCC 3	MFCC 4	MFCC 5	MFCC 6	MFCC 7	MFCC 8	Rough-ness	Zero Cross	Roll off
Attack time	1	.03	-.13	-.01	-.00	.09	-.01	.07	-.12	-.10	.13	.09	-.07	-.08	-.02
Attack slope		1	.03	.12	-.12	-.13	-.05	.07	-.03	.10	.09	-.04	-.20**	.18*	.08
Brightness			1	.12	-.14	-.38**	.28**	-.03	-.07	-.24**	-.04	-.08	-.20**	.59**	.54**
Irregularity				1	.04	-.21**	-.09	-.01	-.05	-.04	.26**	-.03	-.29**	.19**	.06
MFCC 1					1	.00	.00	.00	.00	.00	.00	.00	.12	-.18*	-.22**
MFCC 2						1	.00	.00	.00	.00	.00	.00	.35**	-.30**	-.16*
MFCC 3							1	.00	.00	.00	.00	.00	.09	.16*	.22**
MFCC 4								1	.00	.00	.00	.00	-.01	.00	.01
MFCC 5									1	.00	.00	.00	-.01	-.16*	-.12
MFCC 6										1	.00	.00	.04	-.06	-.14
MFCC 7											1	.00	-.17*	.04	.06
MFCC 8												1	.03	-.09	-.06
Rough-ness													1	-.22**	-.14
Zero - Cross														1	.86**
Roll - off															1

* p <.05, ** p < .01, and *** p < .001.

2. PREDICTIONS

To determine the relationship between the independent variables (acoustic components), and the dependent variables (emotion ratings, instrument ratings, and IADS ratings), principal components analysis was performed, followed by regression analysis. The main goal of the research was to establish whether particular categories of emotion (e.g., happy, sad, anger, fear or disgust; see Ekman, 1992) and timbre are explained by particular acoustic qualities of sound, and to discover how these attributes are related.

Due to the ease with which producers and researchers produce and create many kinds of complex sounds by controlling for specific acoustical properties (Padova et al., 2003), the timbral variations within a single instrument that are used to transmit emotions are more variable and more easily manipulated. In this regard, implications for this research could mean that, if the acoustic feature roughness is found to be a significant predictor for both emotion and timbre in terms of the synthetically created stimuli, that roughness is a main determinant of both timbre and emotion.

Speech perception research has indicated that mel-frequency cepstral coefficients are a major source, or carrier, of information (Loughran et al., 2001. Mfcc’s are the dominant features used for speech recognition (Logan, 2001), and are based on the mel-scale which approximates the human auditory system's response. The mel-scale is based on a mapping between the actual frequency of a sound and its perceived pitch. Due to this underlying relationship between speech and music processing, it is hypothesized that mel-frequency cepstral coefficients will be a significant acoustic component for timbre with instrumental sounds. If mfcc's are also a strong predictor for emotion, it can be foretelling that both emotion and timbre are related in terms of speech sounds. Mfcc's have recently come into the music field as a new point of research; for example, Brown (1998) discriminates between oboe and saxophone sounds by calculating cepstral coefficients.

For the IADS (environmental type sounds) it is not thought that mel-frequency cepstral coefficients will apply in the same way due to the processing used for the different types of sounds. It has been acknowledged that the gap in literature linking the acoustic components of sound and emotion reflects the idea that musical emotions are not like other emotions (Krumhansl, 1997). Environmentally based sounds have real implications for an individual's welfare; these sounds are followed by a bodily reaction, and emotions are essential to physically prepare an individual to perform such an action. Music however does not have such an overt effect; it is hypothesized that though the same acoustic features may not be located for the IADS, it is expected that there will be a connection between the category and emotion within the sounds.

3. EXPERIMENTS 1A AND 1B: INSTRUMENTAL SOUNDS

3.1. Experiment 1a: Instrument Judgment Experiment

Novel stimuli were created to convey particular emotions based on previous research (Gabrielsson & Juslin, 1996; Hailstone et al., 2009; Juslin, 2000; Sloboda, 1991). The stimuli were created, as in Hailstone et al., (2009) to be complex and perceptually distinct to avoid similarities with real musical instruments. This lack of close similarity helped to minimize the effects of learned emotional associations with particular instruments. Synthetic stimuli also removes effects such dynamics or tempo, which may modulate emotional impact (Hailstone et al., 2009).

Participants. A total of 219 participants (73 male, mean age = 18.6, 146 female, mean age = 18.5) participated. Subjects were recruited from the Texas A&M University subject pool and received course credit for participation.

Materials. Stimuli were combinations of two instruments taken from one of four categories of instruments: wind (flute, clarinet, alto saxophone), brass (trumpet, French horn, tuba), string (guitar, piano, violin), and other (bells).

To produce stimuli, ten different instruments were recorded and tuned to approximately 440 Hz. From these ten original sounds, 180 “synthetic” stimuli were created by mixing recordings of two instruments with an audio analysis, editing, and synthesis program (SPEAR, Klingbeil, 2005). Specifically, fast Fourier transform analysis was applied to decompose the sounds into amplitude and frequency components. With the help of laboratory assistants the fundamental frequencies and other frequencies were arbitrarily chosen from each instrument sound and combined to create 180 total stimuli. Laboratory assistants were instructed to take the instrument combinations, for example, frequencies from both flute and clarinet, to create a happy sound using that combination of instruments. For each instrument pair (45 pairs of instruments in all) four sounds were created to sound happy, sad, angry, and fearful. One sound was discarded due to an error in creation leaving a total of 179 sound stimuli.

Procedure. Participants were presented 45 sounds using customized Visual Basic software through Flats stereo headphones. Each stimulus's maximum volume was adjusted and normalized. No participants reported having difficulty hearing the sounds. Stimuli were presented in a random order for each participant. After listening to the stimuli, participants rated each sound on ten different rating scales for instrument type including flute, clarinet, alto saxophone, trumpet, tuba, French horn, violin, guitar, piano, and bell, see Figure 6. These instruments comprised the 179 total stimuli. Participants rated each sound on all ten instruments independently, with each scale ranging from 1 to 7-1 being *strongly disagree* (the degree to which the stimuli, sounded like one of the ten given instruments), and 7 being *strongly agree*, (Figure 6). Results for Experiment 1a will follow the methods for Experiment 1b.

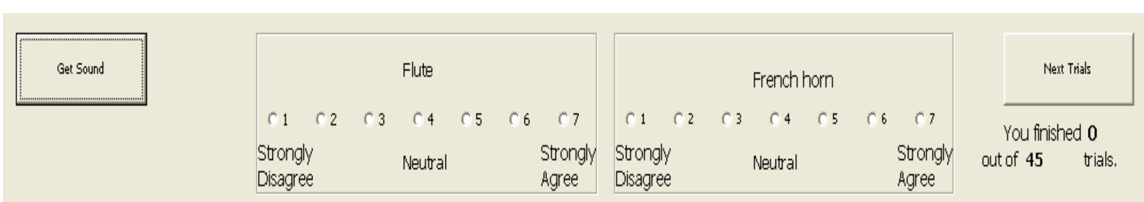


Figure 6. Instrument judgment experiment example. Participants rated each sound on all 10 instruments.

3.2. Experiment 1b. Emotion Judgment Experiment

Having established timbre ratings for synthetically created stimuli, the purpose of this Experiment 1b was to acquire emotion judgment ratings for the same synthetic stimuli.

Participants. A total of 376 participants (202 male, mean age = 19.2 174 female, mean age = 19.2) participated in the experiment for course credit. No participants who participated in Experiment 1a participated in Experiment 1b.

Materials. Stimuli used were the same as Experiment 1a; combination of two instruments taken from one of four categories of instruments; wind (flute, clarinet, alto saxophone), brass (trumpet, French horn, tuba), string (guitar, piano, violin), and other (bells).

Procedure. The procedure of the emotion judgment experiment was identical to that in the Experiment 1a, except for a minor modification. In this experiment, participants were presented 90 sounds, one at a time, and rated each sound on five different rating scales including happy, anger, sad, fear, and disgust. These emotions were chosen based on previous emotion literature (Ekman, 1992). Participants rated each sound on all five emotions; with each emotional scale ranging from 1 to 7-1 being *strongly disagree*, and 7 *strongly agree*, see Figure 7.

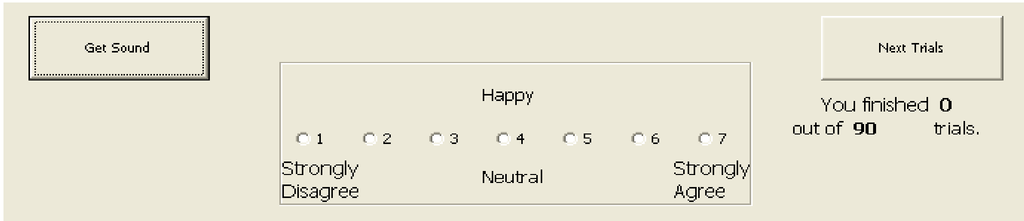


Figure 7. Emotion judgment experiment example. Participants rated each sound on all five emotions.

3.3. Principal Components Analysis (PCA): Experiments 1a and 1b

Using principal components analysis (PCA), a large number of variables are reduced to a smaller, more coherent set of variables. The primary reason for using PCA prior to analyses was to compare responses made for emotion ratings and instrument ratings; PCA allows comparison of the data at a certain percent cutoff of the total variability of the original data (emotion and instrument ratings). This technique works to linearly transform a set of variables into a set of smaller, uncorrelated variables; the goal is to reduce the dimensionality of the original data set (Abdi & Williams, 2010). Because the principal components are uncorrelated, or orthogonal, each one makes an independent contribution to accounting for the variance of the original variables. The first component has the largest possible variance, and explains the largest part of the original data set. The second component is orthogonal to the first component and also works to explain as much of the data from the original data set as possible, and so on for subsequent components.

When measuring two variables, for example, height and weight in a ten hospital patients, it is easy to plot and visualize this data and assess the correlations between the two factors. However, when more than two or three dimensions of data are used, it is difficult to visualize the interactions and correlations within the data set, therefore, PCA is a useful tool to make a large data set more manageable. Figure 8 illustrates this method of dimension reduction used for the dependent variable of instrument ratings in Experiment 1a; the same procedure was applied to emotion ratings for Experiment 1b. The original data in Experiment 1a contained ratings of 179 sounds, for 10 instruments each, and for over 100 participants; a very large data set. PCA works to fit the data into components that account for a certain amount of variance within the data.

A.	Flute	Clarinet	Trumpet	Tuba	Piano	French Horn	Violin	Guitar	Saxophone	Bell
Sound 1	3.43	3.28	2.62	1.83	3.22	2.35	3.01	2.13	2.47	5.8457
Sound 2	3.45	2.64	2	1.47	2.71	1.71	2.50	1.81	2.07	5.96
Sound 3	3.792	2.92	2.28	1.88	2.75	2.16	2.62	2.01	2.50	5.03

B.	PCA 1	PCA 2	PCA 3	PCA 4	PCA 5
Sound 1	2.61	-0.92	-0.22	-0.01	-0.29
Sound 2	1.35	-0.89	-0.29	0.16	-0.05
Sound 3	0.00	-1.29	0.04	-0.41	-0.27

Figure 8. Principal component analysis of instrument ratings. This figure illustrates the method used for PCA to reduce the dimensions of instrument ratings. Figure 8 A shows the original data while Figure 8 B demonstrates the reduction of the original data into principal components. The actual size of original data in Figure 8 part A and B have been decreased by the number of sounds for purposes of explanation.

PCA was used on instrument and emotion responses to reduce the dimensionality of the dependent variables. The cutoff criterion selected, uses the first three components which describe nearly 80% of the variance for the rating data extracted in the timbre judgment and emotion judgment experiments. Methods such as this are based on previous principal components research (Wold, 1987; Abdi & Williams, 2010). See Figure 9 for a visual depiction of percent variance accounted for by each principal component for instrument rating data and Figure 10 for emotion rating data.

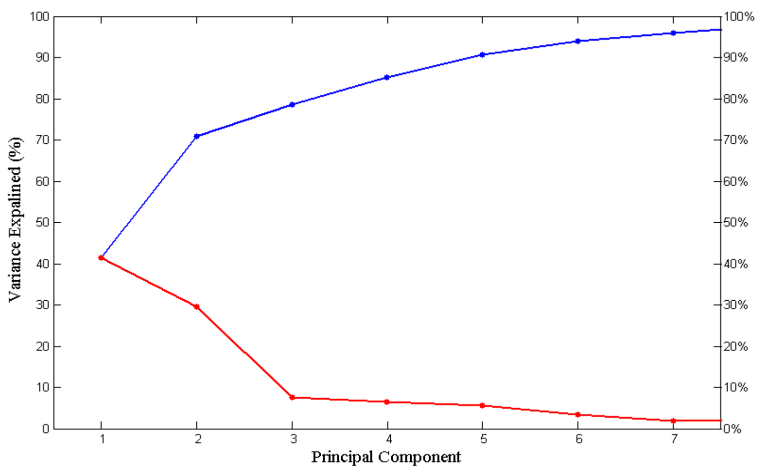


Figure 9. Scree plot of observations for principal components describing instrument ratings. This figure demonstrates the variance in timbre ratings for each principal component of Experiment 1a. Percent variance accounted for by each principal component is indicated by a point on the red line. The blue line indicates cumulative percent variance for the principal components. The first two principal components account for more than 80% of the variance in the instrument rating data.

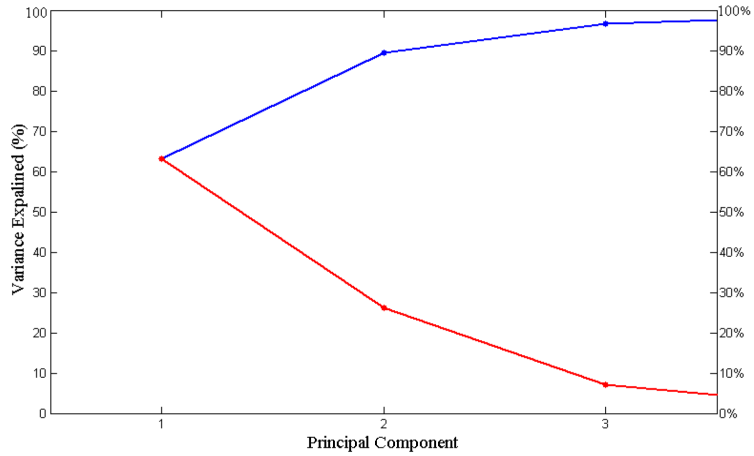


Figure 10. Scree plot of observations for principal components of emotion ratings. This figure shows the variance in emotion ratings for each principal component for Experiment 1b. Percent variance explained for each principal component is specified by a point on the red line, while cumulative percent variance is indicated by the blue line. The first two principal components account for more than 80% of the variance in the emotion rating data.

3.4. Principal Components Analysis (PCA): Experiments 2a and 2b

PCA was also used on IADS category and IADS emotion responses to reduce the dimensionality of the dependent variables. A cutoff criterion for the principal components of 80% of the cumulative percentage of total variation was used.

Three principal components accounted for approximately 80% of the data for the category judgment regression analysis, and three principal components for the emotion judgment regression analysis, see Figure 11 for a visual depiction of percent variance accounted for by each principal component for IADS category rating data, and Figure 12 for IADS emotion rating data.

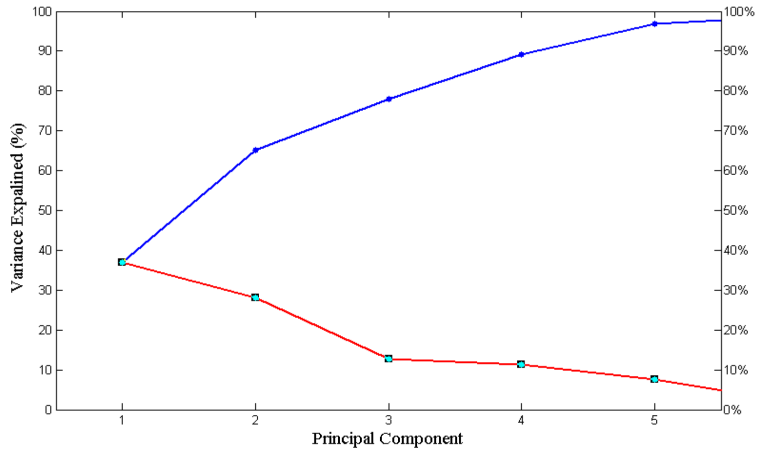


Figure 11. Scree plot of observations for principal components describing IADS category ratings. This figure shows the variance in IADS category ratings for Experiment 2a. The percent variance accounted for by each principal component is noted by a point on the red line, the blue line shows the cumulative percent variance for by the principal components. The first three principal components account for more than 80% of the variance in the IADS category rating data.

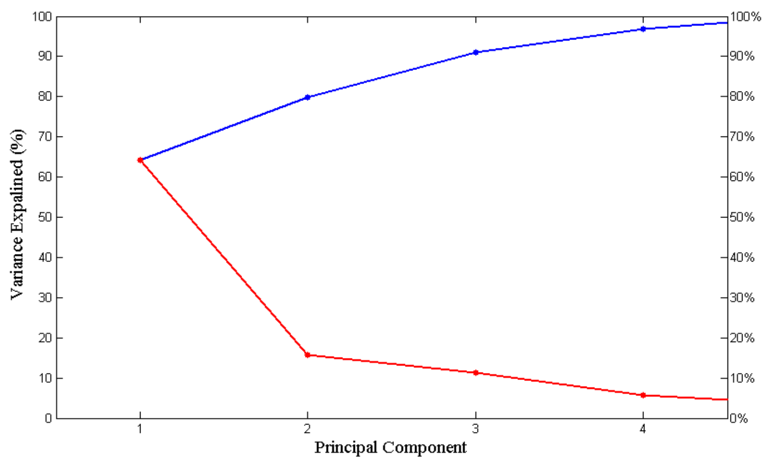


Figure 12. Scree plot of observations for principal components describing IADS emotion ratings. This figure shows the variance in IADS emotion ratings for each principal component of Experiment 2b. The percent variance accounted for by each principal component is indicated by a point on the red line. The blue line shows the cumulative percent variance accounted for by the principal components. The first three principal components account for more than 80% of the variance in the emotion rating data.

3.5. Results. Experiments 1a and 1b

The following sections explain the results of Experiment 1a and 1b. First a preliminary data analysis of Experiment 1a was run to explain general elements of the instrument rating data followed by results of the stepwise regression for Experiment 1a. The same order of presentation is utilized for Experiment 1b.

Stepwise regression analyses evaluate different independent and dependent variables. The first regression analysis uses the independent variable (predictors) and regresses this on the dependent variables (instrument ratings) for synthesized sounds of Experiment 1a. The next stepwise regression is between the predictor variables (independent variables) and emotion ratings of the synthesized sounds from Experiment 1b. The purpose is to locate the acoustic components that can explain both emotion and timbre.

The timbre data alone are able to convey interesting patterns and implications for the results of the Experiments 1a and 1b. Figure 13 shows a preliminary analysis of instrument ratings for the timbre Experiment 1a. From the figure, it is apparent that there is more variability in ratings for the instruments flute, tuba, and bell, over and above the other seven instruments.

It has been noted that the selection of musical instruments is relevant to the expression of emotion in a sound (Balkwill & Thompson, 1999; Gabrielsson, 2001; Gabrielsson & Juslin, 1996; Juslin, 2000). Figure 13 shows the observations for each of the 10 instruments rated for Experiment 1a. From the whiskers of the box plot for the instrument data, it is evident that there is spread within the data. The highest rating for the timbre data did not exceed a value of approximately 6.25, on the scale of 1-7. The median of the ratings for instrument varied between approximately 1.25 and 6.25 signifying some amount of variability within the data. For all 179 sounds rated, most were rated as piano or bell, indicated by the median of the data for piano and bell. The sounds were rated least like the instrument tuba, as the median for this instrument was the lowest for all sounds rated on the ten instruments.

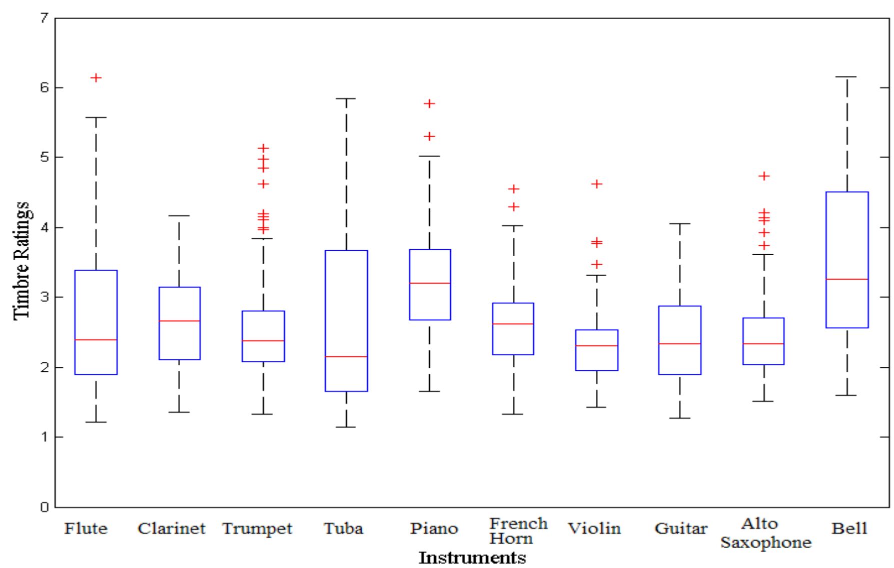


Figure 13. Box plot of observations for timbre ratings. This figure illustrates the timbre ratings for the Experiment 1a. Each box indicates one instrument rated by participants, the median is indicated by the red line in the center of each box, and the edges indicate the 25th and 75th percentiles. The whiskers of each plot indicate the extreme data points, and outliers are plotted outside of the whiskers.

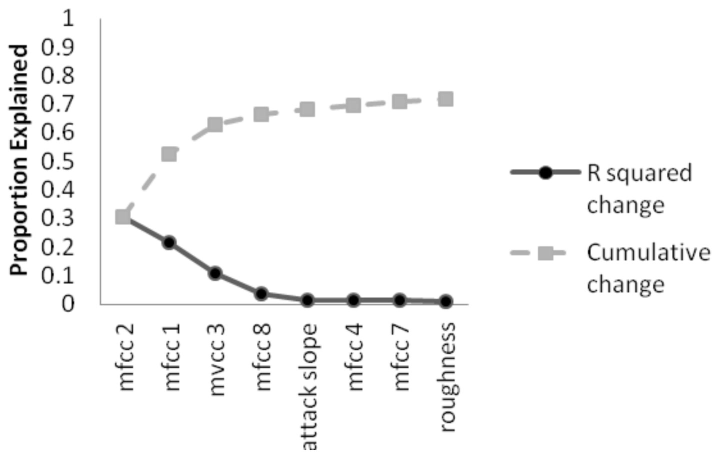
3.6. Results 1a: Instrument Judgment

A step-wise regression was used to determine statistically significant predictor variables; this analysis worked by including predictors step-by-step to the model, to determine which acoustic components could best explains the instrument judgment data.

Principal component analysis was used to reduce the dimensions of the instrument judgment data from ten dimensions to two in the instrument regression for Experiment 1a, which explained 80% of the variance in the instrument rating data. The steepest decline in the data (see Figure 14a and 14b) occurred in the first two components of the instrument judgment data.

For principal component one, the results of this regression indicated that eight acoustic features, out of fourteen total acoustic features could significantly predict instrument ratings, these are as follows; mfcc 2 (β = -.500, $p<.001$), mfcc1 (β = -.441, $p<.001$), mfcc 3 (β = .340, $p<.001$), mfcc8 (β = -.183, $p<.001$), attack slope (β = .123, $p<.01$), mfcc4 (β = -.119, $p<.01$), mfcc 7 (β = -.135, $p<.001$), and roughness 2 (β = -.106, $p<.001$), see Table 2 for R-squared, or percent of variance described by the regression for principal component 1. The R-squared value tells which model works the best to explain the dependent variable, and also conveys the "fit" of the model to the data for each predictor added to the model. Results for principal component two showed that eight acoustic features significantly predicted instrument ratings, these are; mfcc 1 (β = .297, $p<.001$), mfcc 2 (β = -.230, $p<.001$), attack time (β = -.193, $p<.01$), irregularity (β = .114, $p<.05$), mfcc 5 (β = .165, $p<.01$), mfcc 4 (β = -.149, $p<.05$), roll off (β = -.490, $p<.001$), and zero cross (β = .432, $p<.001$), (Table 2).

Figure 14 shows the proportion of R-squared contributed for each addition of a predictor to the model for each principal component, and the proportion of R-squared that was contributed for each addition of a predictor to the model for instrument ratings.



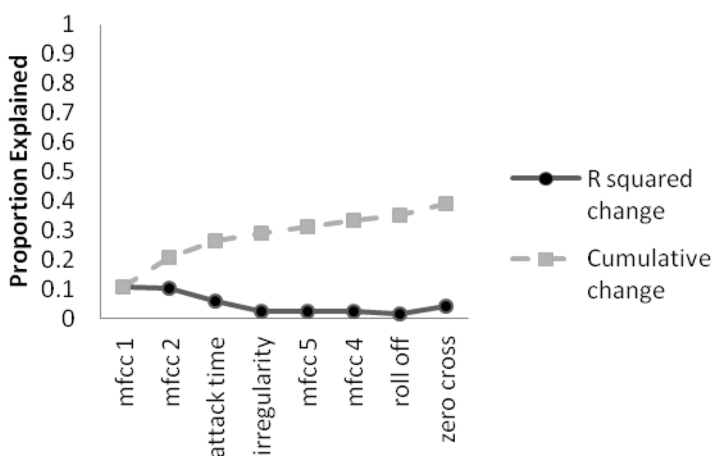


Figure 14. R-squared. Instrument principal component two. This figure illustrates the change in R-squared for each addition of a predictor to the model. The dashed line in each figure demonstrates cumulative change, and the solid line represents the proportion of R-squared for each additional predictor to the model. Figure 14a demonstrates these values of R-squared for principal component one, and Figure 14b shows the values of R-squared for principal component two for the instrument rating data.

Overall, the first two components of instrument PCA work well to describe a majority of the instrument ratings (70.98% of the instrument rating data). The most common features between all of the components are mfcc 1, mfcc 2, and mfcc 4.

Table 2

Significant acoustic components for instrument PCA

	PCA 1	PCA 2
% PCA explained	41.49	29.49
Attack time		X*
Attack slope	X**	
Brightness		
Irregularity		X
MFCC 1	X***	X***
MFCC 2	X***	X***
MFCC 3	X***	
MFCC 4	X**	X**
MFCC 5		X**
MFCC 6		
MFCC 7	X***	
MFCC 8	X***	
Roughness	X***	
Zero Cross		X***
Roll off		X***
R-squared	0.717	0.935

* $p < .05$, ** $p < .01$, and *** $p < .001$.

It is important to note that each principal component is orthogonal from the other; they make an independent contribution in accounting for the variance of the original variables. In the case of the instrument principal components here, this does not seem to hold true due to the many shared predictors between the components.

The results from Experiment 1a, as a whole, show that mfcc 1, mfcc 2, and mfcc 4 are very good predictors of instrument rating data (Figure 14a and 14b; Table 2). To determine if there is a relationship between predictors for timbre and emotion, it is necessary to analyze emotion rating data, where it is expected that mfcc will also be a main contributor to emotion rating data due to the presupposed relationship between timbre and emotion.

A preliminary data analysis for emotion judgments are shown in Figure 15 which depicts observations for each emotion rated in Experiment 1b. From the whiskers of the box plot for the emotion data, it is evident that there is a small amount variation within the data; indicating that perhaps emotion was an easier to access and rate within the sound stimuli. It is also noted that the highest rating for the emotion data did not exceed a value of 6, on the scale of 1-7. The median of the ratings for emotion only varied between approximately 2.8 and 4.0 within the emotion rating data. For all 179 sounds rated, most were rated as fearful, indicated by the median of the data for fear. The sounds were rated least like the emotion happy, as the median for this emotion was the lowest for all sounds rated on the five emotions.

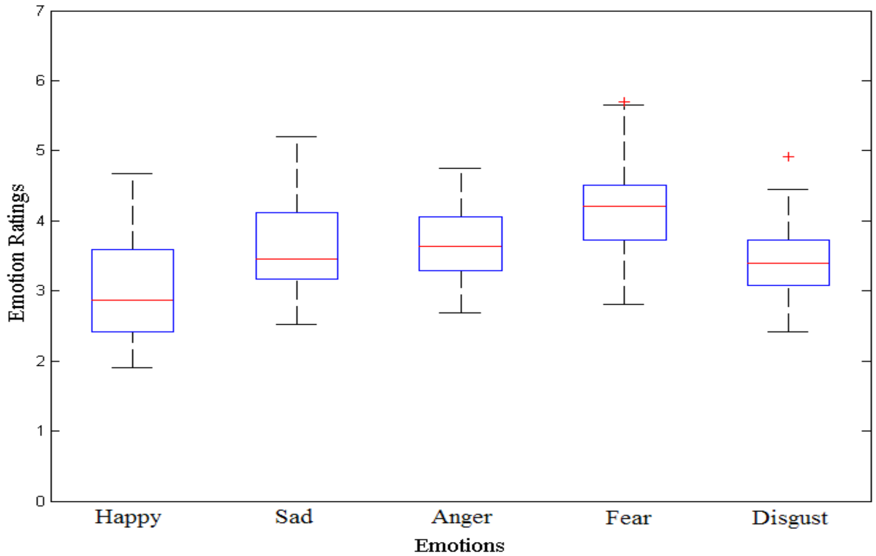


Figure 15. Box plot of observations for emotion ratings. This figure illustrates emotion ratings for Experiment 1b. Each box indicates one emotion rated by participants, the median is indicated by the red line, and the edges show the 25th and 75th percentiles. Whiskers of each plot indicate the extreme data points, and outliers are plotted outside of the whiskers.

3.7. Results 1b: Emotion Judgment

Similarly to the regression for the instrument judgment Experiment 1a, a step-wise regression analysis was used to analyze the collected rating data and acoustic features. Principal component analysis was also used as in Experiment 1a with a cutoff criterion of 80% and a reduction from five to two dimensions.

The results of the regression for emotion ratings of the first principal component indicated three acoustic features significantly predicted emotion ratings; roughness (β = -.517, $p<.001$), mfcc 3 (β = -.184, $p<.01$), and mfcc5 (β = .132, $p<.05$, (Table 3). Five acoustic features of fourteen total significantly predicted emotion ratings for principal component two; mfcc 2 (β = .498, p<.001), mfcc 1 (β = .322, p<.001), mfcc 3 (β = -.296, p<.001), attack time (β = -.157, p<.01), and brightness (β = -.153, p<.01), (Table 3).

Table 3

Significant acoustic components for emotion PCA

Predictors	EPCA 1	EPCA 2
% explained	63.27	26.26
Attack time		X**
Attack slope		
Brightness		X**
Irregularity		
Mfcc 1		X***
Mfcc 2		X***
Mfcc 3	X**	X***
Mfcc 4		
Mfcc 5	X*	
Mfcc 6		
Mfcc 7		
Mfcc 8		
Roughness	X***	
Zero-cross		
Roll off		
R-squared	0.333	0.562

* p <.05, ** p < .01, and *** p < .001.

Figure 16 shows the proportion of R-squared contributed for each addition of a predictor to the model for principal component one from the emotion judgments as well as the proportion of R-squared that was contributed for each addition of a predictor to the model for principal component two from the emotion judgments.

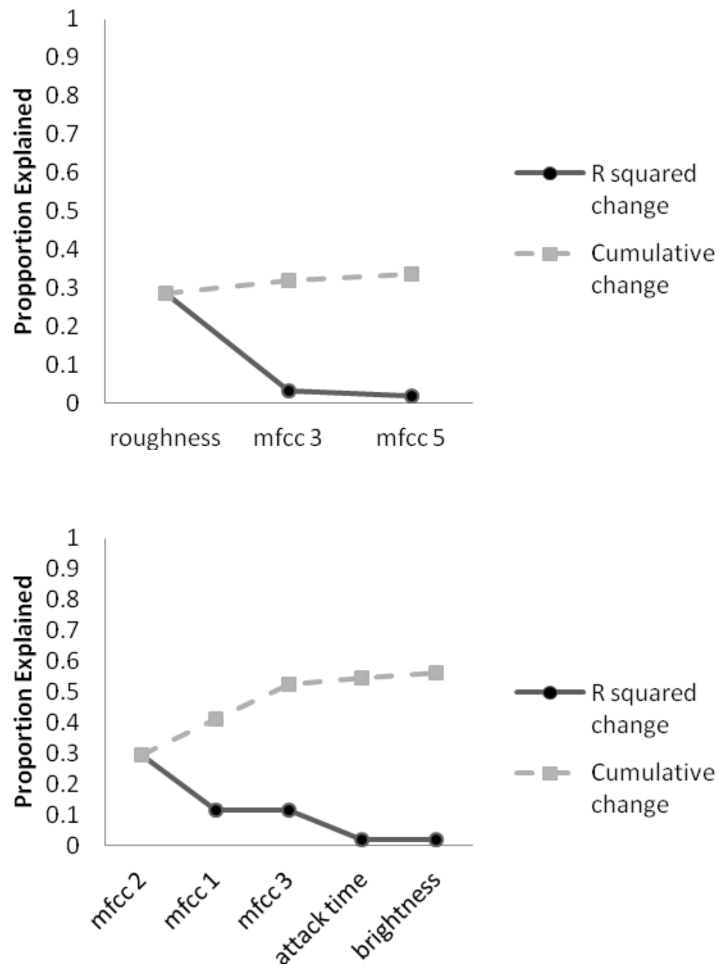


Figure 16. R-squared. Emotion principal component two. This figure illustrates the change in R-squared for each addition of a predictor to the model. The dashed line in each figure demonstrates cumulative change, and the solid line represents the proportion of R-squared for each additional predictor to the model. Figure 16a demonstrates these values of R-squared for principal component one, and Figure 16b shows the values of R-squared for principal component two for the emotion rating data.

In regard to the comparison between the regression results for instrument and emotion, Table 4 lists the shared predictors between the principal components for emotion (EPCA) and timbre (IPCA). It is interesting to note that the predictors, or acoustic components, shared by both timbre and emotion are attack time, mfcc 1-3, mfcc

5, and roughness. Due to the implications of mfcc and speech processing and simulation, this relationship shows that predictors that can explain both emotion and timbre for the synthetically created sounds could also explain speech; though no other predictors were able to do so. This relationship between the synthetic sounds and speech is discussed more in the general discussion section in comparison with the emotion rating and IADS data.

Table 4

Shared predictors for timbre and emotion

Predictors	IPCA 1	IPCA 2	EPCA 1	EPCA 2
% Explained	41.49	29.49	63.27	26.26
Attack time		X*		X*
Attack slope	X*			
Brightness				X*
Irregularity		X*		
Mfcc 1	X*	X*		X*
Mfcc 2	X*	X*		X*
Mfcc 3	X*		X*	X*
Mfcc 4	X*	X*		
Mfcc 5		X*	X*	
Mfcc 6				
Mfcc 7	X			
Mfcc 8	X			
Roughness	X*		X*	
Zero-cross		X*		
Roll off		X*		
R-squared	0.717	0.935	0.333	0.562

* p <.05, ** p < .01, and *** p < .001.

Figure 17 displays the porportion of R-squared for each of the principal components for both the instrument and emotion. This figure represents the percent of the data explained for each principal component. The instrument principal component first explained 41.49% of the isntrument rating data, and principal component two explained 29.49% of the instrument data, or that not accounted for by the first principal component. The primary reason for using PCA is to be able to compare responses made for both emotion and instrument ratings. Figure 17 shows that the difference in percent explained moving from instrument and emotion principal component one, to instrument and emotion principal component two decreases considerably.

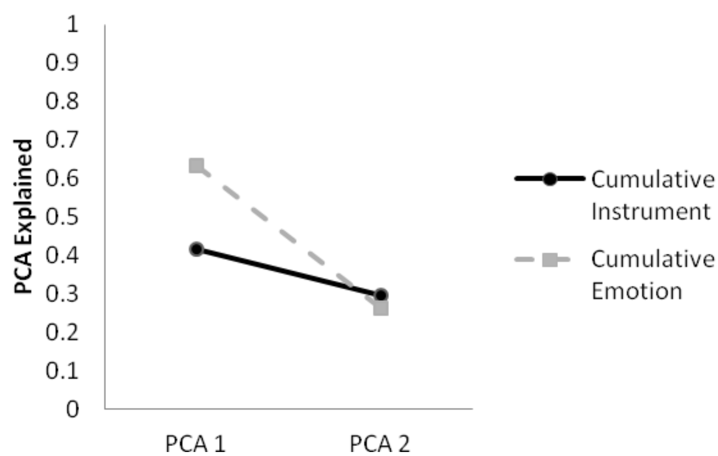


Figure 17. Amount of instrument and emotion rating data explained for each principal component. This figure illustrates the instrument (solid) and emotion (dashed) changes in the percent explained from the first principal component, to the second principal component. This value indicates how much of the instrument data, or emotion rating data, is explained by the principal component.

3.8. Discussion. Experiments 1a and 1b

The conclusions that can be drawn from the results of Experiments 1a and 1b show that timbre components do have an effect on the perception of emotion in sound by normal participants. The shared predictors between emotion and timbre go a long way in answering whether or not the acoustic components can predict both emotion and timbre in sound. Attack time, roughness, and mfcc's were main contributors that could explain both instrument ratings, and emotion ratings. While both roughness and attack time were significant at each stage in the stepwise regression model, (roughness was able to explain much more data for the first principal component than the second), they did not explain the overall instrument or emotion ratings as well as mfcc's (Figure 16a and 16b).

In terms of mfcc's, research by Loughran et al. (2001) found that this particular component was the most useful and efficient predictor to classify musical instruments. Similar findings for acoustic components were observed in Caclin et al., (2005) where it was discovered through the use of multi-dimensional scaling, one major determinant of timbre was attack time. Irregularity was also found to be a salient acoustic feature of timbre. While Caclin utilized timbre dissimilarity ratings, we believe that direct ratings are more effective to understand the implications of timbre and emotion in sound.

Both the instrument and emotion rating data were predicted by very similar acoustic components, mfcc 1 and mfcc 2 were strong predictors for both sets of data. This gives merit to the theory that emotion and timbre are intrinsically related and answers the research question, to what degree or how are these related. One possible determinant of the relationship between timbre and emotion for these instrumental sounds is a possibility of some unique quality embedded in instrumental sounds. This unique quality could extend to type of instrument, possibly woodwind instruments ratings are better predicted by mfcc. It is also possible that there is an intrinsically more interesting quality linking timbre and emotion in terms of instrumental sounds, such a connection could explain why people are so moved by and connected to music.

Overall, the results of this study expand upon other timbre research that has found an explanation of the relationship between timbre and emotion in that particular acoustic features can explain the relationship between timbre and emotion. In this case, for synthetically created instrumental sounds, a relationship was discovered in terms of mfcc.

4. EXPERIMENTS 2A AND 2B: INTERNATIONAL AFFECTIVE DIGITIZED SOUNDS (IADS)

4.1. Experiment 2a. International Affective Digitized Sounds: Category judgment experiment

To further clarify the relationship between emotion, timbre, and sound in a more natural way, it is necessary to use sounds that mimic the environmental world. The IADS include sounds of a cat meowing, carnival noises, human interactions, etc. environmental type sounds. These sounds utilize a simple dimensional view, which "assumes emotion can be defined by a coincidence of values on a number of different strategic dimensions" Bradley & Lang (2007). Dimensional views of emotion have been advocated by a large number of theorists through the years, including Mehrabian and Russell (1974) and Tellegen (1985).

In terms of category rating of sound and emotion, very little research is available. One study by Gygi et al. (2007), had listeners rate145 environmental sounds on 20 semantic dimensions. Intercorrelations of the ratings suggested that 90% of the variance was associated with four factors; harshness, size, complexity, and appeal.

The categories used, power, safe, alive, natural, useful, near, and action, were chosen based on Ekman's (1992) line of work about basic emotions. Basic emotions can be thought of in several ways, first that they are separate and differ in important ways (such as physiology, or behavioral response), this is more the social constructionist view of basic emotions. They can also be viewed in terms of basic meaning that these emotions evolved for adaptive value to deal with fundamental life tasks, or that basic emotions are used for appraisal of a task and they are influenced by ancestral past (Ekman, 1992). In this light, Cosmides, Tooby, & Barkow (1992) focus on the relationship between the structure of psychological mechanisms and human culture (how psychological mechanisms are used to solve adaptive problems). Cosmides, Tooby, & Barkow (1992) look not only at behavioral descriptions of brain function, but also information-processing - the how and why information processing has the functional properties it does. These functions are adaptive problems such as finding a mate, finding food, avoiding predation etc., which is why the categories of power, safe, alive, natural, useful, near, and action were chosen.

The purpose of this experiment is to gain a better understanding of the relationship between sound and emotion in terms of acoustic features, and to see whether the same features will be used to predict emotion and categories with non-instrumental sounds.

Participants. A total of 361 participants (185 male, mean age = 18.6, 176 female, mean age = 18.5) participated in the experiment for course credit.

Materials. Stimuli used were the International Affective Digitized Sounds, Stevenson & James (2008).

Procedure. The procedure of the emotion judgment experiment was identical to that in the emotion, and instrument judgment experiment (Experiments 1a and 1b) with minor modifications. In this experiment, participants were presented 106 sounds, one at a time, and rated each sound on seven different rating scales including power, safe, alive, natural, useful, near, and action. These categories were chosen based on previous music and evolution literature (Balkwill et al., 2004; Hailstone et al., 2009). Participants rated each sound on all five categories; with each category scale ranging from 1 to 7-1 being *strongly disagree*, and 7 *strongly agree*, see Figure 18.

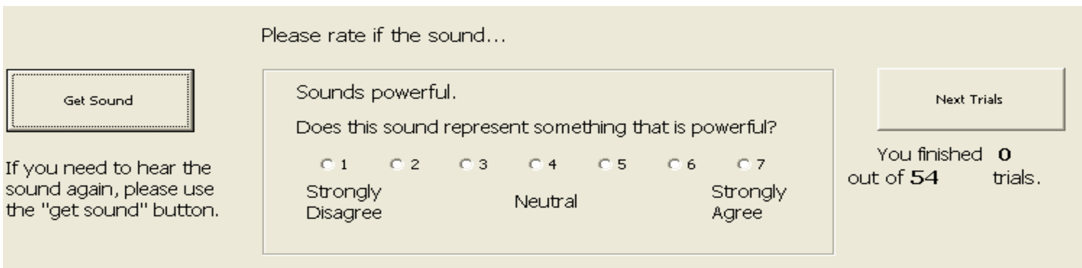


Figure 18. IADS judgment experiment example. Participants rated each sound for all 7 categories.

4.2. Experiment 2b. International Affective Digitized Sounds: Emotion Judgment Experiment

Previously collected data from Stevenson & James (2008) were analyzed for this experiment. Five sounds used in Stevenson & James (2008) were not included in our analysis due to exclusion in the collected IADS data.

Participants. College students, both female and male, attending Introductory Psychology classes at the University of Florida participated as part of a course requirement. At least 100 participants rated each sound of which approximately half were female, a total of 167 sounds were rated.

Materials. Stimuli used were the International Affective Digitized Sounds, Bradley & Lang (2007).

Sixty sounds were obtained from a variety of formats ~e.g., CDROM collections, audiotapes, recordings made in the laboratory using actors and actresses from the University of Florida’s Theatre department, and digitized. Each sound was edited to a 6 seconds. Peak sound intensity at presentation ranged from 64 to 81 dB as measured using a Quest 1700 Precision Impulse Sound Level Meter, and varied according to natural volumes in the environment. Rise and fall times varied across stimuli, and were controlled to prevent eliciting startle responses. Presentation of sounds was controlled and each sound was presented for 6 s over a pair of JBL 4311 Control Monitor speakers. Ratings for each sound were completed using the Self-Assessment Manikin (Lang, 1980). SAM ranges from a smiling, happy figure to a frowning, unhappy figure, representing the pleasure dimension, and SAM ranges from an excited, wide-eyed figure to a relaxed, sleepy figure for the arousal dimension. For the dominance dimension, SAM ranges from a large figure (in control) to a small figure (dominated).

Procedure. Procedures to collect ratings were from Stevenson & James (2008).

Participants were presented 111 sounds from the IADS using MATLAB 5.2 (Math Works, Inc., Natick, MA) software with the Psychophysics Toolbox extensions (Brainard, 1997; Pelli, 1997) running on a Macintosh computer, through Beyerdynamic DT 100 headphones. Each stimulus’s maximum RMS was adjusted to 1 and presented at full volume. Stimuli were presented in a random order for each participant. Following the sound, participants saw a series of five rating scales including happiness, anger, sadness, fear, and disgust, with scales presented in random orders. These emotions were chosen for two reasons: their inclusion in nearly all discrete categorical theories of emotion, and their inclusion in databases of facial expression. Participants rated each sound on all five emotions independently, with each discrete emotional scale ranging from 1 to 9—1 being *not at all* and 9 being *extremely*. Participants had one hour to complete all ratings. In the case that participants did not finish within one hour, only sounds for which all five ratings had been given were scored, resulting in 71–75 scores for each sound (*M* 5 73.7). No participants reported having any difficulty hearing the sounds.

4.3. Results. Experiments 2a and 2b

The following sections feature the results of Experiment 2a and 2b. First a preliminary data analysis of Experiment 2a to explain general features of the IADS category rating data is detailed, and then the results of the stepwise regression for Experiment 2a. After presenting the results of Experiment 2a the preliminary data analysis of Experiment 2b and the results of the stepwise regression analysis with the IADS emotion rating data of Experiment 2b are presented.

The stepwise regression analyses use the independent variable of predictors (acoustic components) and regresses these upon the dependent variable of either IADS category or IADS emotion ratings, respectively. The purpose is to locate the acoustic components that can explain both emotion and timbre, in other words, whether acoustic qualities of sound can predict particular categories of emotion (e.g., happy, sad, anger, fear or disgust) or categories, and how such attributes are related.

Figure 19 shows a preliminary data analysis of the overall category ratings for the IADS category data. Categories were chosen based off of work by Gygi et al., (2007) as well as

Ekman's (1992) work on basic emotions. For the IADS, the highest rated category was action. From the whiskers of the box plot for the category data, it is evident that there is some variation within the data. The median of the ratings, as indicated by the horizontal red line in each box, varied between approximately 3.5 and 5.75 signifying a moderate amount of variability within the data. For all 106 sounds rated, most were rated as belonging to the category action, indicated by the median of the data for the category action. The use of this category, action, shows listeners' use of adaptive functioning, according to Tooby & Cosmides (1989), as well as Ekman (1992) as a major determinant of this category. Tooby & Cosmides (1989) state that an evolutionarily derived task analysis can help to produce a hypothesis about the structure of human cognitive processes and by understanding the environmental sounds in terms of adaptive categories, it is easier to understand their use evolutionarily. The sounds were rated least belonging to the category natural, as the median for this instrument was the lowest for all sounds rated on the seven categories.

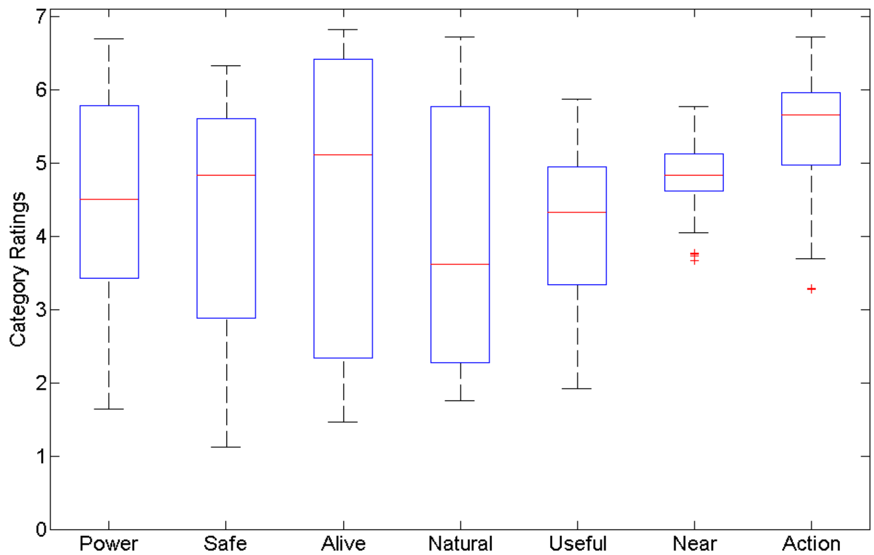


Figure 19. Box plot of observations for IADS category ratings. Each box indicates one category rated by participants, the median is indicated by the red line in the center of each box, and the edges indicate the 25th and 75th percentiles. The whiskers of each plot indicate the extreme data points, and outliers are plotted outside of the whiskers.

4.4. Results 2a: IADS Category Judgments

Categorization of sounds is influenced by factors such as goals, variability, and theories (Barsalou, 1991; Fried & Holyoak, 1984; Murphy & Medin, 1985). Through using multidimensional scaling, Gygi et al, (2007) found that perceived similarities among environmental sounds are strongly determined by the acoustic features of those sounds, such as harmonicity, spectral spread, continuity, periodicity, and envelope modulation.

Principal component analysis was used, as in Experiment 1a, to reduce the dimensions of the category judgment data from seven different dimensions to three, with a cutoff criterion of approximately 80% of the data being explained within these selected dimensions

The results indicated that only one acoustic feature, zero cross, significantly predicted IADS category rating for principal component one (β = -.225, $p<.05$), (Table 5), for principal component two, roll off was the only acoustic feature significantly predicted IADS category ratings (β = .441, $p<.01$), (Table 5) and irregularity was the only reliable acoustic feature for principal component three (β = -.210, $p<.05$), (Table 5).

Overall these acoustic components did not work well to explain the category rating data as indicated by the R-squared values (Table 5). The best predictor, roll off, explained 20% while other predictors explained less than 10% of the principal component data.

Table 5

Significant acoustic components for IADS category PCA

	IADS PCA 1	IADS PCA 2	IADS PCA 3
% PCA explained	36.99	28.20	12.72
Attack time			
Attack slope			
Brightness			
Irregularity			X*
MFCC 1			
Roughness			
Zero Cross	X*		
Roll off		X***	
R-squared	0.051	0.194	0.044

* p <.05, ** p < .01, and *** p < .001.

A preliminary data analysis of the emotion rating data is shown in Figure 20. From the whiskers of the box plot for the IADS emotion data it is evident that there is very little variation within the data; there is even less variation than that of the emotion ratings for synthesized sound in Experiment 1a, indicating that perhaps emotion was a more difficult entity to rate for the IADS stimuli. The median of the ratings for emotion only varied between a little bit under and a little bit over 2.0, signifying a small, limited variability within the IADS emotion rating data.

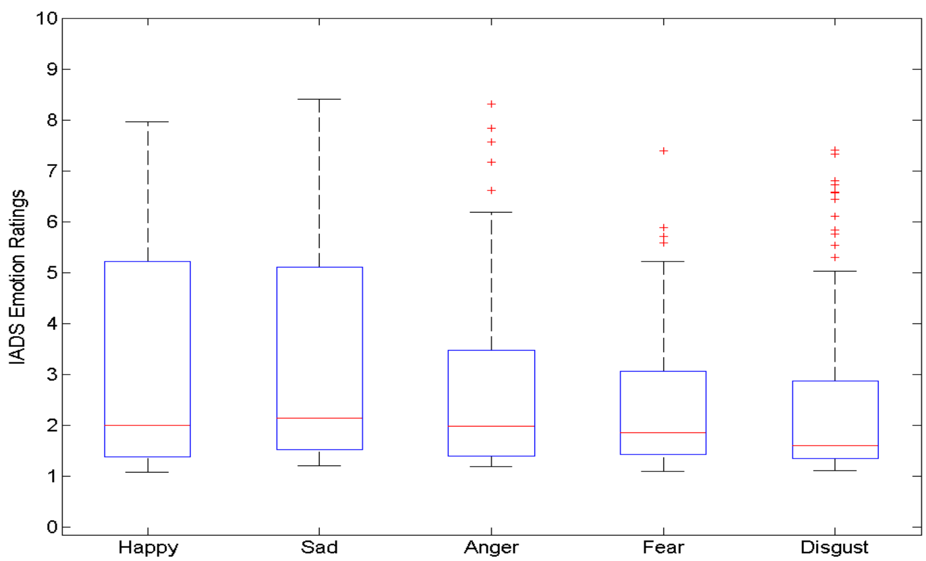


Figure 20. Box plot of observations for IADS emotion ratings. Each box indicates one emotion rated by participants, the median is indicated by the red line in the center of each box, and the edges indicate the 25th and 75th percentiles. The whiskers of each plot indicate the extreme data points, and outliers are plotted outside of the whiskers.

4.5. Results. Experiment 2b: IADS Emotion Judgments

The goal of Experiment 2b was to determine if acoustic components map onto listeners' emotion ratings for environmental sounds in the same way that they map on to the synthesized sounds created and used in Experiments 1a and 1b. As in Experiments 1a and 1b principal components analysis and step-wise regression were utilized to analyze the rating data.

The first principal component had no significant acoustic features that could predict IADS emotion ratings. Results for principal component two found two significant acoustic features; brightness (β = .458, $p<.01$), and zero cross (β = -.340, $p<.01$). Results for the regression on the third principal component indicated that two significant acoustic features; roll off (β = -.447, $p<.01$), and brightness (β = .302, $p<.05$), (Table 6).

The best overall model for the emotion data is indicated in Table 6, in terms of R-squared. Figure 21 shows the proportion of R-squared contributed for each addition of a predictor to the model for principal component two from the IADS emotion judgments, as well as the proportion of R-squared contributed for each addition of a predictor to the model for principal component three from the IADS emotion judgments.

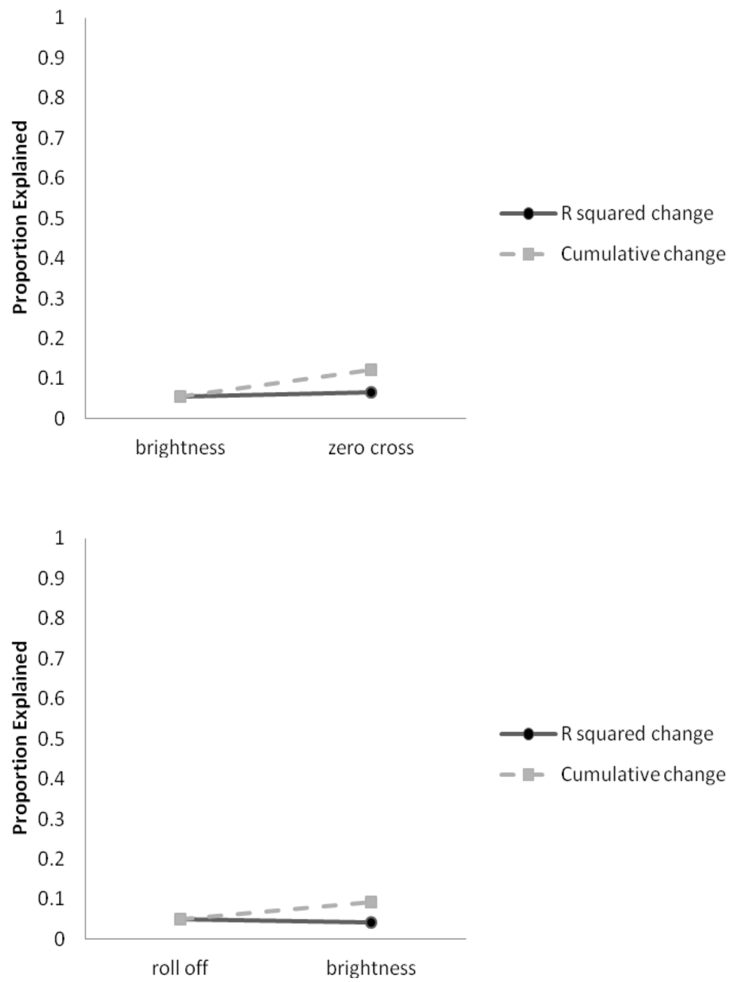


Figure 21. R-squared. Emotion principal component three. This figure illustrates the change in R-squared for each addition of a predictor to the model. The dashed line in each figure demonstrates cumulative change, and the solid line represents the value of R-squared for each additional predictor to the model. No component could explain emotion for principal component one. Figure 23a demonstrates change in amount of R-squared in the model for principal component two for the emotion rating data. Figure 23b demonstrates change in amount of R-squared in the model for principal component three for the emotion rating data.

These results suggest that there is little overlap between acoustic features that can explain the IADS emotion in regard to principal components 1, 2 and 3; however, the two principal components may describe separate features of emotion in sound.

Table 6

Significant acoustic components for IADS emotion PCA

	IADS PCA 1	IADS PCA 2	IADS PCA 3
% PCA explained	64.17	15.64	11.23
Attack time			
Attack slope			
Brightness		X***	X*
Irregularity			
MFCC 1			
Roughness			
Zero Cross		X**	
Roll off			X**
R-squared	-	.122	.094

* p <.05, ** p < .01, and *** p < .001.

Matching predictors for IADS category (CPCA) and IADS emotion (EPCA) data are shown in Table 7; respective p-values are indicated by an asterisk. It is interesting to note that the predictors, or acoustic components, shared by both category and emotion are zero-cross and roll off. These are very different components than those found for Experiments 1a and 1b. This reveals that the link between the predictors used to explain timbre and emotion for the synthetically created instrumental stimuli and that of category and emotion for IADS stimuli are different.

Table 7

Matching acoustic components for IADS category PCA (CPCA) and IADS emotion PCA (EPCA)

Predictors	CPCA 1	CPCA 2	CPCA 3	EPCA 1	EPCA 2	EPCA 3
% PCA explained	36.99	28.20	12.72	64.17	15.64	11.23
Attack time						
Attack slope						
Brightness					X***	X*
Irregularity			X*			
MFCC 1						
Roughness						
Zero Cross	X*				X**	
Roll off		X***				X**
R-squared	0.051	0.194	0.044	-	.122	.094

* p <.05, ** p < .01, and *** p < .001.

Figure 22 displays the percent of the data explained by each principal component for both the IADS category and IADS emotion data. The first principal component for IADS category explained 36.99% of the category rating data, and the second explained 28.2%, and the third explained 12.72% of the category data not accounted for by either principal component one or two. The first principal component for the IADS emotion explained 64.17% of the emotion rating data, the second component described 15.64% of the emotion rating data, and the third component described 11.23% of the emotion rating data not accounting for data already explained by principal components one and two.

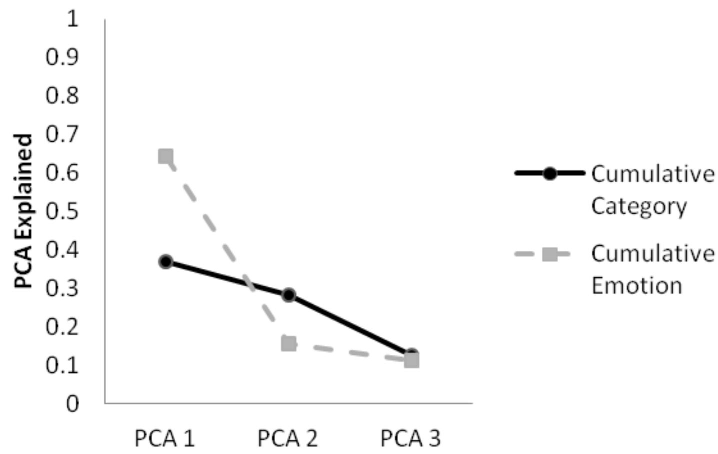


Figure 22. Amount of category and emotion rating data explained for each principal component This figure illustrates the category (solid) and emotion (dashed) changes in the percent explained from the first principal component, to the second principal component. This value of percent explained tells how much of the category or emotion rating data respectively, is explained by the principal component.

4.6. Discussion. Experiments 2a and 2b

The results of Experiments 2a and 2b suggest that timbre does not have an effect on the perception of emotion in sound by normal participants; this is due to the small number of shared predictors for the IADS category, and IADS emotion rating data. In Experiments 1a and 1b, mfcc's were found to be a main contributor of explaining both instrument ratings, and emotion ratings, however, this was not the case for Experiments 2a or 2b. Though Juslin & Laukka (2001, 2003) were able to locate timbral properties that could express affect and some basic emotions relevant to both real and synthetic instruments, this idea according to the results for Experiments 2a and 2b does not apply to the IADS. There was not a significant link between timbre and emotion for the IADS sounds.

One possible reason for the deficit in the relationship between timbre and emotion for these IADS is the small variation within the sound stimuli, or within the ratings for the sounds. Perhaps listeners do not feel the same emotional response from the IADS as from instrumental sounds. This leads to a goal for future research to find the acoustic components that do link IADS, or environmental type sounds in terms of timbre and emotion, perhaps there are acoustic components that can better explain emotion within environmental sounds.

Overall, shared predictors between category and emotion for the IADS sounds of the PCA components show that there is a weak relationship compared to timbre and emotion ratings for the synthetically created stimuli of Experiments 1a and 1b. The results of this study do not work well to explain the relationship between category and emotion.

5. GENERAL DISCUSSION

5.1 Overview

The goal of this research was to determine how timbre and emotion are related in terms of acoustic components in synthetically created instrumental sound stimuli and IADS stimuli. In Experiment 1a participants took part in an instrument judgment task, where the objective was to determine the timbre of synthetic sound stimuli. In Experiment 1b participants rated the same synthetic sound stimuli and performed an emotion judgment task to identify the emotion of the sound stimuli. As hypothesized, the results show that mel-frequency cepstral coefficients were largely responsible for both timbre and emotion. Experiment 2a utilized previously collected emotion rating data of the International affective digitized sounds (Bradley & Lang, 2007). In Experiment 2b participants performed a category rating task on the same IADS. As hypothesized it was found that there was not a significant strong relationship between the acoustic components, timbre, or sound for the IADS. Taken together, these two experiments show that there is a perceptual difference between the relationship of timbre and emotion for instrumental sounds and IADS.

5.2. Implications

Hailstone et al., (2009) claim that it is the timbre of a sound that affects perception of emotion in music. The underlying difference in function of the timbres for instrumental sounds and IADS possibly creates a division in the way they are processed. This research assists in clearing up the poorly defined relationship between perceptual characteristics of a sound and the emotion information they express.

The importance of acoustic features that convey emotion in music and sound has been observed by many studies (Caclin et al., 2006; and Hailstone et al., 2009), but few have been able to make specific conclusions regarding individual acoustic features. This study illustrated that specific acoustic features of timbre, such as mel-frequency cepstral coefficients, could predict both emotion and instrument judgments for non-environmental (instrumental only) stimuli.

Results found in this study are consistent with previous evidence on the effects of timbre on emotion judgments (Balkwill & Thompson, 1999; Hailstone et al., 2009) and suggest that there is some overlap between acoustic features that explain both emotion and instrument judgments.

Though this research does not endeavor to ultimately answer or refute the tension that exists between the body of researchers that attend to a functional difference between music and language, and that which finds evidence for shared features, it can at least shed new light in the field. In taking the view that both language and music share functional processes, we can further research the origin and function of music and language.

5.3. Mel-frequency Cepstral Coefficients

The most dominant acoustic feature that explained both emotion and instrument judgments (Experiments 1a and 1b) were Mel-frequency cepstral coefficients. Mfcc's are features that describe the spectral shape of a sound and are used in speech recognition software, music classification, and audio classification research. The successful use of mfcc's in speech recognition is due to its ability to represent the amplitude spectrum of speech. In the current study, mfcc's are significant for determining emotion and instrument judgments, this could be due to the lack of temporal information in the synthesized sounds; no rhythm or beat information was included and no competing information was present within the sound signal.

Mfcc was a significant acoustic component in Experiment 1, however the same components were not found for Experiment 2. Why was mfcc found to be a significant acoustic predictor for the instrumental sounds? The instrumental stimuli were able to provide a link between timbre and emotion. This finding is in accordance with past research stating a known general link between timbre and emotion (Hailstone et al., 2009). With the inclusion of this research, there has been further evidence for the identification of a specific acoustic component of sound relating both timbre and emotion for instrumental sounds. This component, mfcc, is largely used in speech recognition and music recognition software and production. The fact that mfcc's were found as a good predictor of timbre and emotion suggests that there is an underlying relationship between speech sounds and instrumental sounds with regard to emotion.

Past research by Dolgin and Adelson (1990) tested whether acoustic features of emotional speech are parallel to the emotion in music. They did this by composing musical pieces with varying articulation (such as staccato, and legato), varying tempo (allegro, moderato, largo), and motion (step, skip). Findings showed above-chance accuracy as early as four years of age. This gives a good indication that the emotional associations to music, aside from beginning at an early age, also map on to facial and vocal expressions of emotion (Krumhansl, 1997). The results of this study are consistent with previous evidence on the effects of timbre on emotion judgments (Balkwill & Thompson, 1999; Hailstone et al., 2009), insofar as instrumental sounds are concerned.

Mfcc's were found relevant to the link between timbre and emotion for instrumental sounds; however they were not uncovered as a significant acoustic predictor for the IADS. The lack of congruency between the two experiments could have been due to the variation of timbre for different emotions. Hailstone et al. (2009) much like Juslin & Laukka, (2001; 2003), found that properties of timbre could express affective valences related to both real and synthetic instruments, including some basic emotions. The current experiments, however, did not find this to be the case for these IADS stimuli. For example, fear may be difficult to convey using purely timbral cues without regard to dynamic variations or tempo (Gabrielsson & Juslin, 1996; Sloboda & O'Neill, 2001).

It is possible that timbre can explain emotion for only instrumental and not IADS. A solution to answer this issue might be to limit IADS to a particular category or type of sound, possibly more related to instrumental sounds to see if the same effect is acquired.

It is plausible to think that listeners may make judgments differently for different types of stimuli. Instruments' timbres are comprised of and relate to the type of instrument, as well as properties of that particular instrument. For example, a violin has a different timbre than a flute because it is a string instrument, it is created from wood and not metal, which contributes to the differing timbres. The IADS do not contain the same type of timbre information; they are related to evolutionary goals such as that of safety, and power. These sound stimuli encompassed adaptive problems such as finding a mate, finding food, avoiding predation etc., which is why the categories of power, safe, alive, natural, useful, near, and action were chosen.

5.4. Sound, Speech and Evolution

Previous research suggests that we process different types of sounds in different ways. Perceiving timbre is presumed to rely upon the capacity to perceive and process differences such as the difference between musical instruments, or voices. These differences are fundamental to everyday human functioning and timbre analysis is a fundamental task of the auditory system (McAdams & Bigand, 1993; Godyke et al., 2003). Research using instrumental sound stimuli has found that the influence of instrument timbre on emotion may apply not only to instrumental sounds, but to the processing of other types of sounds in different contexts. However, mfcc was used in the speech processing and was also found as a predictor for the processing instrumental sounds. Due to this connection it can be speculated that the way in which instrumental sounds are processed is directly related to the processing of speech.

Evolutionarily, it has been argued that the brain mechanisms for processing timbre in music evolved for the representation and evaluation of vocal sounds (Juslin & Laukka, 2003). Research in this domain has argued that musical timbres might share acoustic components with emotional vocal expressions (Juslin & Laukka, 2001, 2003); the findings from Experiment 1a and 1b confirm this notion. Features of timbre, such as attack, or low or high frequencies may be able to indicate a form of emotion in music, for example, a "dull" spectral quality is associated with sadness in music, whereas the “brash” quality conferred by prominent high frequencies is associated with anger (Juslin & Laukka, 2001). These may generalize to the expression of emotion through other structural cues in music, the expression of vocal emotion, or expression of emotion in other modalities such as gesture (Hailstone et al., 2009; Sloboda & Juslin, 2001).

5.5. Music background and sound perception

Musical background has been shown to effect the perception of sound. It has been shown that listeners’ understanding of emotion in music is affected by their familiarity with the tonal system (e.g., Western music) and by their sensitivity to basic perceptual cues (Balkwill & Thompson, 1999). No data were collected from participants to indicate involvement or level of musical experience or expertise. The combination of musical knowledge information with instrumental music knowledge may have had an effect on emotion and timbre ratings. For example, it could be determined that mfcc’s are stronger predictors of timbre and emotion for those participants with a higher musical knowledge, or that mfcc’s are a strong predictor despite previous musical knowledge.

6. CONCLUSIONS

This research attempted to link timbre, sound and emotion in terms of acoustic cues. The same acoustic cues could explain emotion people infer from a sound and instrument people identified of the same sound. The results imply that perception of emotion in sound as well as judgment of instrument identity is related to timbre. Specifically, we suggest that the shared acoustic cues are the element of timbre that influences emotion judgment.

REFERENCES

Abdi, H., and Williams, L. (**2010**). “Principal component analysis,” Wiley Interdisciplinary Reviews: Computational Statistics **2**, 433-459.

Balkwill, L.-L., & Thompson, W. F. (**1999**). “A crosscultural investigation of the perception of emotion in music: Psychophysical and cultural cues,” *Music Perception,* **17**, 43–64.

Balkwill, L. L., Thompson, W. F., and & Matsunaga, R. (**2004**). “Recognition of emotion in Japanese, Western, and Hindustani music by Japanese listeners,” JPN. Psychol. Res. **46**, 337–349.

Barsalou, L.W. (**1991**). “Deriving categories to achieve goals,” in *The Psychology of Learning and Motivation: Advances in Research and Theory*, edited by G.H. Bower. (Academic Press, New York), pp. 1-64.

Berger, K., (**1964**). ”Some factors in the recognition of timbre,” J. Acoust. Soc. Am. **36**, 1888-1891.

Bradley, M. M., & Lang, P. J. (**1999**). International affective digitized sounds (IADS): Stimuli, instruction manual and affective ratings (Tech. Rep. No. B-2). Gainesville, FL: The Center for Research in Psychophysiology, University of Florida

Bradley, M. M. & Lang, P. J. (**2007**). “The international affective digitized sounds (2nd Edition; IADS-2): Affective ratings of sounds and instruction manual,” Technical report B-3. University of Florida, Gainesville, Fl.

Brainard, D. (**1997**). “The Psychophyics Toolbox”. Spatial Vision **10**, 433-436.

Bregman, A., Liao, C., Levitan, R. (**1990**). “Auditory grouping based on fundamental frequency and formant peak frequency,” Can. J. Psychol. **44**, 400-413.

Brown, J (**1998**). “Computer identification of musical instruments using pattern recognition with cepstral coefficients as features,” J. Acoust. Soc. Am. **105**, 1933-1941.

Caclin, A., McAdams, S., Smith, B., & Winsberg, S. (**2005**). “Acoustic correlates of timbre space dimensions: A confirmatory study using synthetic tones,” J. Acoust. Soc. Am. **118**, 471-482.

Caclin, A., Brattico, E., Tervaniemi, M., Naataned, R., Morlet, D., Giard, M.H., & McAdams, S. (**2006**). “Separate neural processing of timbre dimensions in auditory sensory memory,” J. Cognitive Neurosci. **18**, 1959-1972.

Chartrand, J.P., and Belin, P. (**2008**). “Superior voice timbre processing in musicians,” Neurosci. Lett. **405**, 164-167.

Cosmides, L., Tooby, J., and Barkow, J. (**1992**). *Introduction: Evolutionary psychology and conceptual integration, the Adaptive Mind* (Oxford University Press, New York), pp. 3-18

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

Dolgin, K. G., & Adelson, E. H. (**1990**). “Age changes in the ability to interpret affect in sung and instrumentally presented melodies,” Psychol. Music **18**, 87-98.

Ekman, P. (**1992**). “Are there basic emotions?” Psychol. Rev. **99**, 550-553.

Fried, L., and Holyoak, K. (**1984**). “Induction of category distributions: A framework for classification learning,” J. Exp. Psychology Learn. **10**, 234-257.

Gabrielsson, A., & Juslin, P. N. (**1996**). “Emotional expression in music performance: Between the performer’s intention and the listener’s experience,” Psychol. Music **24**, 68–91.

Gabrielsson, A. (2001). “Emotions in strong experiences with music,” in *Music and Emotion: Theory and Research,* edited by P.N. Juslin, and J.A. Sloboda. (Oxford University Press, Oxford), pp. 431–452.

Grey, J. M., and Moorer, J. A. (**1977**). “Perceptual Evaluations of Synthesized Musical Instrument Tones”. J. Acoust. Soc. Am. **62** 454- 62.

Godyke, K.N., Altenmuller, E., Moller, J., & Munte, T., (**2003**). “Change in emotional tone and instrumental timbre are reflected by the mismatch negativity,” Cognitive Brain Res. **21** 351-359.

Gygi B., Kidd, G., and Watson, C. (**2007**). “Similarity and categorization of environmental sounds,” Percept. Psychophys. **69**, 839-855.

Hailstone, Julia C., Omar, Rohani, Henley, Susie M. D., Frost, Chris, Kenward, Michael G. and Warren, Jason D. (**2009**) “It's not what you play, it's how you play it: Timbre affects perception of emotion in music,” Q. J. Exp. Psychol. **62**, 2141-2155.

Hajda, J. M., Kendall, R., Carterette, E., & Harshberger, M. (**1997**). “Methodological issues in timbre research,” Percept. Cog. Music **12**, 253-306.

Helmholtz, Hermann von (**1885**). *On the Sensations of Tone as a Physiological Basis for the Theory of Music*, Second English Edition, translated by Alexander J. Ellis. (Longmans, Green, and Co., London).

Hevner, K. (**1935**). “Expression in music: A discussion of experimental studies and theories,” Psychol. Rev. **42**, 186–204.

Johnson-Laird, P. & Oatley, K. (**1992**) "Basic emotions, rationality, and folk theory," Cognition Emotion **6**, 201-23.

Johnstone, T., and Scherer, K. R. (**2000**). “Vocal communication of emotion,” in, *Handbook of Emotions,* edited by I. T. Jolliffe (Guilford Press, New York), pp. 112.

Juslin, P.N. (**2000**). “Cue Utilization in Communication of Emotion in Music Performance: Relating Performance to Perception,” J. Exp. Psychol. **26**, 1797-1813.

Juslin, P. N., & Laukka, P. (**2001**). Impact of intended emotion intensity on cue utilization and decoding accuracy in vocal expression of emotion. Emotion, 1, 381–412.

Juslin, P.N. and Laukka (**2003**). “Communication of emotions in vocal expression and music performance: Different channels, same code?” Psychol. Bull. **129**, 770-814.

Klingbeil, M. (**2005**). “Software for spectral analysis, editing, and synthesis,” in Proceedings of the ICMC, Barcelona, Spain, pp. 107–110.

Koelsch, S., (**2005**). “Neural substrates of processing syntax and semantics in music,” Curr. Opin. Neurobiol. **15**, 207-212.

Krumhansl, C.L. (**1989**). “Why is musical timbre so hard to understand?” in, *Structure and perception of electroacoustic sound and music*, edited by S. Nielzen, and O. Olsson, (Excerpta Medica, Amsterdam), pp. 43-53.

Krumhansl, C.L. (**1997**). “An exploratory study of musical emotions and psychophysiology,” Can. J. Exp. Psychol. **51**, 336-352.

Ladd D.R., Silverman K.A., Tolkmitt F., Bergmann G., Scherer K.R. (**1985**). “Evidence for the independent function of intonation contour type, voice quality, and F0 range in signaling speaker affect,” J. Acoust. Soc. Am. **78**, 435-44.

Lartillot, O., Toiviainen, P., Eerola, T. (**2008**) “A Matlab Toolbox for Music Information Retrieval”, in *Data Analysis, Machine Learning and Applications, Studies in Classification, Data Analysis, and Knowledge Organization*, edited by C. Preisach, H. Burkhardt, L. Schmidt-Thieme, and R. Decker, (Springer, New York), pp 261-268.

Logan, B. (**2001**). “Mel frequency cepstral coefficients for music modeling,” in Proceedings of the ISMIR, Plymouth, Massachusetts.

Loughran, R., Walker, J., O’Neill, M., O’Farrell, M. (**2001**). “The Use of Mel-frequency Cepstral Coefficients in Musical Instrument Identification,” in Proceedings of the ICMC Finland, pp. 1825–1828.

McAdams, S., Cunible, J.C., (**1992**). “Perception of timbral analogies,” Philos. Trans. R. Soc. Lond., B Biol. Sci., **9**, 336-383.

McAdams, S., Bigand, E. (**1993**). *Thinking in sound: The Cognitive Psychology of Human Audition* (Clarendon Press, Oxford).

McAdams, S., Winsberg, S., Donnadieu, S., Soete, G. D., Krimphoff, J. (**1995**). "Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes," Psychol. Res. **58**, 177-192.

Mehrabian and Russell (**1974**). *An approach to environmental psychology* (MIT Press, Cambridge).

Miller, J.R., & Carterette, E.C. (**1975**). "Perceptual space for musical structures," J. Acoust. Soc. Am. **58**, 711-720.

Murphy, G., and Medin, D. (**1985**). "The role of theories in conceptual coherence," Psychol. Rev. **92**, 289 – 316.

Padova, A, Bianchini, L., Lupone, M., & Balardinelli, M., (**2003**). "Influence of specific spectral variations of musical timbre on emotions in the listeners," in Proceedings of the 5th Triennial ESCOM Conference, Germany, pp 227-230.

Patel, A., (**2009**). *Music, language, and the brain* (Oxford University Press, New York).

Pelli, D. G. (1997). "The Videotoolbox software for visual psychophysics: Transforming numbers into movies." Spatial Vision **10**, 437-442.

Plomp R, & Levelt WJ. (**1965**). "Tonal consonance and critical bandwidth," *J. Acoust. Soc. Am.* **38,** 548–60

Plomp, R. (**1970**). "Timbre as a multidimensional attribute of complex tones," in *Frequency Analysis and Periodicity Detection in Hearing*, edited by R. Plomp and G. F Smoorenburg (A. W. Sijthoff, Leiden).

Rasch, R.A., & Plomp, R. (**1982**). "The perception of musical tones," in *The Psychology of Music*, edited by D. Deutsch (Academic Press: New York), pp. 1-13.

Risset, J. C. & Wessel, D. L. (**1982**). “Exploration of timbre by analysis and synthesis,” in *The Psychology of Music*, edited by D. Deutsch (New York: Academic).

Samson, S., Zatorre, R., & Ramsay, J. (**1997**). “Multidimensional scaling of synthetic musical timbre: perception of spectral and temporal characteristics,” Can. J. Exp. Psychol. **51**, 307-315.

Scherer, K., (**1995**). “Expression of emotion in voice and music,” J. Voice **9**, 235-248.

Scherer, K. (**2003**). “Vocal communication of emotion: A review of research paradigms,” Speech Commun. **40**, 227-256.

Sloboda, J.A. (**1991**). “Music structure and emotional response: Some empirical findings,” Psychol. Music **19**, 110-120.

Sloboda, J.A., Juslin, P., 2001. “Psychological perspectives on music and emotion,” in *Music and Emotion: Theory and Research*, edited by P. N. Juslin, and J. A. Sloboda (Oxford University Press, Oxford) pp. 71– 104.

Sloboda, J. A., & O’Neill, S. A. (**2001**). “Emotions in everyday listening to music,” in *Music and Emotion: Theory and Research*, edited by P. N. Juslin & J. A. Sloboda (Oxford University Press, Oxford) pp. 415-430.

Stevenson, R., and James, T. (**2008**). “Affective auditory stimuli: Characterization of the International Affective Digitized Sounds (IADS) by discrete emotional categories,” Behav. Res. Meth. **40**, 315-321.

Tellegen, A. (**1985**). “Structures of mood and personality and their relevance to assessing anxiety, with an emphasis on self-report,” in *Anxiety and the anxiety disorders*, edited by A. H. Tuma & J. D. Maser (Erlbaum, New Jersey) pp 681-706.

Tooby, J., & Cosmides, L. (**1989**). “Evolutionary psychology and the generation of culture, part I, Theoretical considerations,” Ethol. Sociobiol. **10**, 29-49.

Tooby, J., and Cosmides, L. (**1990**). “On the universality of human nature and the uniqueness of the individual: The role of genetics and adaptation,” J. Pers. **58**, 17-67.

Wedin, L., and Goude, G. (**1972**). "Dimension analysis of the perception of instrumental timbre," Scand. J. Psychol. **13**, 228-240.

Wessel, D.L., & Grey, J.M. (**1978**). “Conceptual structures for the representation of musical material,” in Proceedings of the IRCAM Technical Report No. 14, Paris, France.

Wold, S., Esbensen, K., & Geladi, P. (**1987**). “Principal Component Analysis,” Chemometr. Intel. Lab. **2**, 38-52.

VITA

Name: Casady Diane Bowman

Address: Texas A&M University, 4235 TAMU, College Station, TX 77843

Email Address: casadyb@tamu.edu

Education: B.A., Psychology, Oklahoma State University, 2007
M.S., Cognitive Psychology, Texas A&M University, 2011