2009-12-14

Acoustic Mediation of Vocalized Emotion Identification: Do Decoders Identify Emotions Idiographically or Nomothetically?

Michael Kenneth Lauritzen
Brigham Young University - Provo

Acoustic Mediation of Vocalized Emotion Identification: Do Decoders Identify Emotions Idiographically or Nomothetically?

Michael K. Lauritzen

A dissertation submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Bruce L. Brown, Chair
Matthew P. Spackman
Robert D. Ridge
Sam A. Hardy
Ross Flom

Department of Psychology

Brigham Young University

December 2009

ABSTRACT

Acoustic Mediation of Vocalized Emotion Identification: Do Decoders Identify Emotions Idiographically or Nomothetically?

Michael K. Lauritzen

Department of Psychology

Doctor of Philosophy

Most research investigating vocal expressions of emotion has focused on one or more of three questions: whether there exist unique acoustic profiles of individual encoded emotions, whether the nature of emotion expression is universal across cultures, and how accurately decoders can identify expressed emotions. This dissertation begins to answer a fourth question, whether there exist unique patterns in the types of acoustic properties persons focus on to identify vocalized emotions. Three hypotheses were tested: first, whether acoustic patterns are interpreted idiographically or nomothetically as reflected in a comparison of individual vs. group lens model identification ratios; second, whether there exists a decoder by emotion interaction for scores of accuracy; and third, whether such an interaction is mediated by the acoustic properties of the vocalized emotions. Results from hypothesis one indicate there is no difference between individual and group identification ratios, demonstrating that vocalized emotions are decoded nomothetically. Results from hypothesis two indicate there is not a significant decoder by emotion interaction on scores of accuracy, demonstrating that decoders who are generally good (or bad) at identifying some vocalized emotions tend to be generally good (or bad) at identifying all vocalized emotions. There are, however, significant main effects for both emotion and decoder. Anger and happiness are more accurately decoded than fear and sadness. Perhaps most importantly, multivariate results from hypothesis three indicate strong and consistent differences across the four emotions in the way they are identified acoustically. Specifically, decoders identify anger by primarily focusing on spectral characteristics, fear by primarily focusing on frequency (F0), happiness by primarily focusing on rate, and sadness by focusing on both intensity and rate. These acoustic mediation differences across the emotions are also shown to be nomothetic, that is, they are surprisingly consistent across decoders.

Keywords: Emotion, Vocal, Decoding, Acoustic, Properties, Identification

ACKNOWLEDGEMENTS

Thank you to my entire committee for your questions, critiques, and suggestions. Most importantly, thank you to Dr. Bruce Brown, without whom the present research would certainly never have been completed. I cannot imagine there is another advisor on this campus who is more giving of his time, knowledge, and energy to help a student than are you. The many late nights (or, perhaps more accurately, early mornings) we spent together in the lab will be cherished memories for the rest of my life.

Thanks to my parents also, who provided me with the moral and spiritual background necessary to succeed not only academically, but more importantly, in everyday life. To them I owe the deepest of debts for their endless love, sacrifice, and support.

And finally, thanks to Tara, who, more than anyone else, deserves (not to mention wants) acknowledgement on this page. It is for you that I have devoted myself to this endeavor. You kept me on course when my desire waned. You are my best friend, closest confidant, and greatest motivation. You are the reason I find joy in what I do. Thanks, Wife, I love you.

Table of Contents

List of Tables

List of Figures

Acoustic Mediation of Vocalized Emotion Identification: Do Decoders[1] Identify Emotions Idiographically or Nomothetically?

Research in the area of emotion expression has historically focused on two types of expression, facial and vocal. Of these two, facial expression has been the subject of a majority of the research in modern psychology (see for example, Ekman, 1972; Ekman, Sorenson, & Friesen, 1969; Elfenbein, Marsh, & Ambady, 2002; Izard, 1980). This is perhaps because of the emphasis given by Charles Darwin to facial expressions in his book, *Expression of the Emotions in Man and Animals* (1872). Not only was this book perhaps the first comprehensive investigation into the facial expression of emotions but it continues to be among the more influential. Darwin noted that most species of animals capable of expressing emotion did so in similar ways. For example, he observed that dogs, cats, monkeys, and humans all express anger by pulling back their lips to bear their teeth, furrowing their brows, and so forth. As will be seen below, this Darwinian tradition is still prevalent in contemporary research on emotion expression.

Although vocal expressions of emotion have not received the same degree of attention as have facial expressions, philosophers have for centuries recognized that effective emotional expression is heavily reliant upon particular characteristics of the voice. For example, Cicero, the famed orator of ancient Rome, suggests in his *de Oratore* that a clear, strong voice is “the one thing above all others that sets off and supports a speaker’s eloquence” (Cicero, 2001 pg. 102).

[1] Banse and Scherer (1996) proposed the use of the terms *decoding* and *encoding* rather than *receiving* and *sending* or other similar terminology because, the authors suggest, these suggested terms capture both the underlying processes associated with expressing and interpreting emotions as well as the research methods employed to investigate them. The terms should not be confused as to suggest the existence of a particular “code” of emotional communication.

Perhaps the earliest explicit description of the acoustic properties associated with vocal expressions of emotion, though, appears several centuries prior to Cicero. Aristotle, likely not coincidentally one of Cicero's primary philosophical influences, wrote the following in his *Nicomachean Ethics* in the mid- to late-fourth century B.C.:

> And it depends on the voice, as to how we ought to manage it in reference to each several passion; when, for instance, we should employ a loud, when a low, and when a moderate pitch of voice; and on the manner in which we should employ its tones, viz. The acute, the grave, and the intermediate; and on certain rhythms in reference to each; for the points, in reference to which they conduct their inquiries, are three, viz. the loudness of the voice, the fitness of its tones, and its rhythm. (Aristotle & Hobbes, 1890, pg. 204)

Of special interest in the quote above are the three properties that Aristotle found important in differentiating between expressions of emotions: loudness, tonal properties, and rhythm. As will be discussed in more detail below, these same properties (albeit labeled slightly differently) are still used by researchers of vocalized emotion today.

Although Aristotle's description was among the first recorded general statements about the acoustic nature of vocal emotion expression, the earliest extensive analysis of sounds associated with emotions comes from a familiar source: Darwin (1872). In the same volume made famous for Darwin's descriptions of facial and gestural expressions of emotion, Darwin also details some of the vocal properties of emotion expressions. For example, the primary expression for joy in both humans and other animals, according to Darwin, is a vocal expression. In humans we call this expression laughter. In other beings, it might be called something else but Darwin suggests that it is usually vocalized and incited similarly. Apes, for example, "…likewise utter a reiterated sound, corresponding with our laughter, when they are tickled, especially under the armpits" (Darwin, 1872 p.199).

Since Darwin, research on emotion expression in general has focused on three questions: whether there exist unique expression profiles for individual emotions, whether these profiles can be interpreted universally across culture, and the degree to which persons can accurately identify others' emotions through their expressions (see Banse & Scherer, 1996, for a discussion of these questions specific to vocal emotion expressions). This dissertation will begin to address a fourth question regarding the patterns in acoustic properties used by perceivers to decode others' vocal expressions of emotion. Before discussing this fourth question explicitly, however, it is important to have a better understanding of how it is affected by the first three.

Emotion Profiles

The concept of profiling emotions is intriguing to both psychologists and laypersons alike. The lay public's interest in profiling emotion expressions is evidenced in the recent success of Fox Television's hit series 'Lie to Me'. An excerpt of the program description taken from the show's website is below:

> Dr. Cal Lightman (Tim Roth) is the world's leading deception expert. If you lie to Lightman, he'll see it in your face and your posture or hear it in your voice. If you shrug your shoulder, rotate your hand or even just slightly raise your lower lip, Lightman will spot the lie. By analyzing facial expressions and involuntary body language, he can read feelings ranging from hidden resentment to sexual attraction and jealousy. (http://www.fox.com/lietome/).

As the program description indicates, interpretation of emotional expressions through body language and voice can make for compelling T.V. drama. Dr. Cal Lightman, fictional deception expert extraordinaire, always knows what a criminal is really feeling. And he is never wrong. Compelling as the prospect of profiling emotions is, the question exists whether such accurate profiling is merely Hollywood fiction or if it is in fact possible in real life.

Profiling Methodologies. Most studies attempting to identify unique acoustic profiles for particular emotions are similar in their methodological approach. However, they do tend to differ from each other on some key aspects, including whether such studies use trained or untrained encoders and how vocalized emotions are elicited. Of the most relevance to the current study is deciding upon a method for eliciting vocalized emotions.

Researchers interested in identifying vocal profiles of emotion typically utilize one of three methods for eliciting emotion. They may employ recordings from real life emotionally charged situations; they may artificially induce particular emotions in speakers; or, they may utilize simulated or acted emotions. There are pros and cons to each of these methodologies. Utilizing real life recordings provides high mundane realism but it typically does not allow for accurate analysis of the acoustic properties because the quality of the recordings tends to be quite low. Utilizing artificially induced emotions provides the researcher with a great deal of control and typically produces comparable voice samples across participants (see Scherer, 2003). However, it cannot be assumed that experimentally inducing emotions produces the same emotional states for all participants. In addition, experimentally induced emotions typically produce only relatively weak affect, making it more difficult to differentiate among emotional expressions (see Scherer, 2003). For these reasons, most researchers tend to prefer using simulated emotions. Although it can appropriately be argued that simulated emotions tend in some sense to be caricatures of actual emotions, in that they tend to overemphasize some key elements of the emotion, this is not necessarily problematic because, as Scherer (1986) argues, it tends to be these key elements that differentiate particular emotions from each other.

It should be noted that some studies utilize a combination of these methodologies. For example, Flom and Bahrick (2007) first induced emotions in their encoders using the Facial Action Coding System as proposed by Ekman and Friesen (1978), and then instructed them to simulate the respective emotions. Although this combined approach can be effective, it did not seem appropriate for the current study because one question of interest (not addressed in the current paper but addressed in another study utilizing the same encoded emotion data, Spackman, Brown, and Otto, 2009) involved the degree to which trained versus untrained actors differed in their expressions. Therefore, "coaching" the participants to in some sense improve their expressions could have reduced the expected variance between trained and untrained encoders. For this reason, Spackman et al did not combine methodologies.

Evidence For and Against Acoustic Profiles of Emotion. There seems to be fairly general consensus in the literature on facial expressions of emotion that there exist unique facial profiles for individual emotions (see e.g., Ekman, 1972; 1994; Izard, 1971). Of greater interest to the current study, however, is whether there exist unique acoustic profiles for vocalized emotions.

In 1986, Klaus Scherer presented a model for future research of vocal affect expression. In his model he hypothesized that individual emotions would differ from each other on a number of different acoustic parameters including various elements of fundamental frequency (often referred to as F0 and experienced by the listener as pitch), formants, intensity, rate, and spectral noise. Scherer's hypotheses have been influential in much of the subsequent research attempting to identify unique acoustic profiles of vocalized emotion. Although not all of Scherer's hypotheses have held throughout the years, research continues to suggest that the acoustic properties of speech rate, voice intensity, voice quality (or timbre), and F0 are among the most powerful cues in terms of their effects on listeners' ratings of emotional expressions (see, e.g., Juslin, 1997c, 2000; Juslin & Madison, 1999; Lieberman & Michaels, 1962; Scherer & Oshinsky, 1977).

Results from most studies suggest that acoustic patterns associated with particular emotions tend to be quite distinct. This is evidenced in Pittam and Scherer's (1993) review of the early decades of research on five encoded vocalized emotions: anger, fear, sadness, joy, and disgust. The authors concluded that anger tends to be characterized by an increase in pitch and mean energy; fear characterized by increased pitch, range of pitch, high-frequency energy, and rate; sadness characterized by decreased pitch, range of pitch, and mean intensity; and joy characterized by increased pitch, range of pitch, variability of pitch, and mean energy. Disgust, however, tended to have quite inconsistent results in the literature.

The degree to which these differences constitute unique, individual profiles, however, still appears to be in question. In a more recent review, Juslin and Laukka (2003) conducted a meta-analysis of 104 vocal emotion studies. Among the more notable findings in this analysis was that emotions generally referred to as positive such as happiness and tenderness tend to show more regularity in pitch, rate, and intensity than do negative emotions such as anger or sadness. Otherwise, results seemed to generally lend support to those from the Pittam and Scherer (1993) review and Scherer's (1986) hypotheses. However, the authors admit the results from their meta-analysis were not as clear cut as they had hoped, but suggest this is likely due to the inconsistent nature of statistical reporting in the literature rather than inconsistent patterns in vocal emotion expression.

In an attempt to broaden the scope of emotions for which profiles might be identified, Banse and Scherer (1996) conducted a study of 14 vocalized emotions. They found that, after removing variance accounted for by speaker gender and idiosyncratic speaker differences, emotions still accounted for a majority of the variance in acoustic properties (see p. 623). In addition, the study provided fairly compelling evidence for individual differences in intensity and valence or quality aspects between emotions. However, no specific profiles for these emotions were presented.

In sum, many studies have attempted to identify unique profiles for individual emotions. Some studies are more convincing than others, but overall there seems to be some consensus about what the vocal profiles of certain emotions might look like. However, recent work from Spackman, Brown, and Otto (2009) suggests there still remain some questions about the methodological approaches employed in these studies. For example, some researchers assume that, by using speakers trained in emotion expression, they are somehow getting at a purer form of the emotion and therefore can get a better picture of the emotion's profile. Results from Spackman, et al.'s (2009) study, however, suggest that the degree of training has little to no effect on acoustic profiles (see Spackman, et al, 2009 for a more in depth discussion on other methodological concerns in the vocal emotion research literature).

In addition to raising methodological concerns about the literature, results from the Spackman, et al. (2009) study also cast some doubt on the existence of distinct individual acoustic profiles for particular vocalized emotions. Instead of single emotions having singular profiles, the results from their study suggest, it appears as though there might exist multiple identifiable profiles for any given emotion.

Implications on Decoding Patterns. The implications of the nature of acoustic profiles for encoded emotions on the patterns of such profiles for decoded emotions are of central interest to the current study. If, as results from the Spackman, et al (2009) study suggest, there exist multiple modes of expressing emotion, it makes some intuitive sense that persons would employ varied methods to interpret or decode emotions. This relationship between encoded and decoded emotions seems to be an important issue that, to this author's knowledge, has not been addressed directly in the literature.

Universality

Not entirely distinct from the research attempting to identify unique expression profiles for emotion is the research suggesting that such profiles are universal. Although the universality of emotion expressions will not be addressed in the current research, a discussion of the nature of emotion expression research would be incomplete without discussing universality. As alluded to above, one of the primary elements of Darwin's theory was that emotional expressions are universal. That is, there exists for each emotion a distinct facial and vocal profile regardless of culture and, at least to some extent, regardless of species.

Darwin's suggestion has, at least historically (see Russell, 1994, for a detailed criticism of this perspective), received a great deal of support in the literature. For example, one study (Ekman, et al., 1987) compared interpretations of Caucasians' facial emotion expressions made by persons from nine countries. The results, summarized in Table 1, suggest a high degree of agreement across the different cultures.

Table 1

Percentage of Ekman et al.'s subjects in each culture who correctly identified the portrayed emotion

Nation	Happiness	Surprise	Sadness	Fear	Disgust	Anger
Estonia	90	94	86	91	71	67
Germany	93	87	83	86	61	71
Greece	93	91	80	74	77	77
Hong Kong	92	91	91	84	65	73
Italy	97	92	81	82	89	72
Japan	90	94	87	65	60	67
Scotland	98	88	86	86	79	84
Sumatra	69	78	91	70	70	70
Turkey	87	90	76	76	74	79
United States	95	92	92	84	86	81

For decades the results of similar studies went seemingly unchallenged, until Russell (1994) famously took issue with the methodology and interpretation of results in a review of the facial expression literature. He suggested that Ekman and others overemphasized the evidence supporting universality and neglected to identify cross-cultural differences. What ensued became one of the more legendary debates in the history of social psychology. Ekman (1994) rebutted Russell's arguments, suggesting that Russell set up a straw man of universality, one in which 100% agreement across all expression decoders within and across cultures was necessary. In his rebuttal, Ekman suggested there is room for some cultural (and language) differences in interpretation but that, overall, people tend to successfully identify facial expressions from members of other cultures. Of course, Russell retorted back suggesting that he recognized the existence of what he called "minimal universality" but did not back down from his claims that Ekman and others had misinterpreted their results (see Russell, 1995). The debate has still not been settled but it is fair to say that the majority of researchers of facial emotion expressions today presume the existence of at least some degree of universality in expressions.

Unlike research in the area of facial expression of emotion, the issue of universality has not received a great deal of attention in the vocal expressions literature. However, Juslin and Laukka (2003), in a review of over 100 studies of vocal emotion expression argued there is enough evidence to conclude that vocal expressions are "based on innate, fairly stable, and universal affect programs" (p. 775). The authors based this conclusion on evidence from several sources in both the vocal and facial expression research. These sources include evidence that children born deaf and blind express emotions in mostly the same way as their typically developed peers (Eibl-Eibesfeldt, 1973), evidence from cross-cultural studies of facial expression and emotional intelligence (Elfenbein, Marsh, & Ambady, 2002), and evidence for the existence of discrete categories of emotions called basic emotions (de Gelder, Teunisse, & Benson, 1997; de Gelder & Vroomen, 1996; Etcoff & Magee, 1992).

Unfortunately, although Juslin and Laukka (2003) argue these sources of evidence are directly applicable to vocal emotion expression, most of the studies they cite did not directly investigate vocal expressions. Instead, Juslin and Laukka draw inferences about vocal expressions from other types of expressions to make their point. To be fair, they did conclude that findings from vocal expression studies alone did indicate that vocal expressions of emotion can be accurately identified cross-culturally. However, the accuracy was much lower than for within-cultural vocal expression. Therefore, more research will be necessary to determine whether there exist universal patterns in vocal emotion expression.

Implications on the Universality of Decoding Patterns. If there do in fact exist universal, evolutionary based patterns in vocal expressions of emotion, it would likely also be the case that there exist universals in emotion interpretation. Although this question will not be addressed in the present research, it is a question that warrants future investigation.

Accuracy

The issue of universality has yet to be entirely resolved but what appears to be much clearer is that people in general are quite good at identifying emotions. Studies in which participants are asked to identify emotions from content-free speech samples typically show between a 55% to 65% accuracy rate (see, e.g., Banse & Scherer, 1996; Murray & Arnott, 1993; Pittam & Scherer, 1993; Scherer, 1986; Scherer, Banse, Wallbott, & Goldbeck, 1991; Van Bezooijan, 1984; Wallbott & Scherer, 1986). Although these numbers might not seem impressive at first, Scherer (1989) suggests that an accuracy rate of 60% in many of these studies is approximately five times more accurate than would be expected by chance alone. Of the four most frequently studied emotions (anger, fear, happiness, and sadness), anger and fear typically result in the highest levels of accuracy, followed by sadness and then happiness.

Research isolating facial expressions of emotion generally result in higher accuracy levels across all emotions, typically around 75% (see, Ekman, 1972; 1994 for reviews). There are several potential reasons for this difference. One is that there might be wider variability in vocal expressions of similar emotions than there is in facial expressions. For example, the differences between facial expressions of glee and enjoyment are not as great as the differences between the vocal expressions of these two emotions. Similarly, a second reason persons may have more difficulty identifying vocalized emotions than they do facial expressions is physiological in nature. The number movements humans can make with their faces are limited by their musculature, whereas the kinds of noises a person can make, although limited, are much more likely to vary greatly (see Scherer, 2003 p. 236 for discussion surrounding the dynamic nature of vocal stimuli and Ekman, 1972 for discussion of the limited number of basic facial muscle configurations).

Accuracy levels of both vocal and facial expression identification should be interpreted with some caution, however. Critics argue there are a number of problems with assessing identification accuracy scores. Their primary argument deals with the issue of forced response options. In the majority of vocal emotion expression studies, including the research presented in this paper, respondents are instructed to identify the vocalized emotion from a set number of responses (usually about five). It is duly noted that such a methodology has the potential to artificially inflate accuracy levels. One rebuttal to this concern, however, is that, as discussed above, studies employing a forced choice model typically report that emotions are identified at much higher than chance levels (typically around five times greater than chance, according to Scherer, 1989). Because these levels are so high, it is still likely that raters would identify vocalized emotions at greater than chance levels even if the number of response options was not limited. There is simply a lot of room for accuracy rates to decline before they are no longer deemed significantly better than chance.

Implications on Decoding Patterns. It could be argued that, because people are, in general, quite good at accurately identifying vocalized emotions, they all decode them in a similar fashion. However, most studies only report accuracy levels for their entire sample and not for individual decoders. For this reason, it is incorrect to assume that, simply because vocalized emotions are accurately identified in the aggregate, individual decoders all identify them with equal accuracy. Perhaps more importantly, regardless of whether individual decoders are similar in the accuracy of their ratings, it should not simply be assumed that they all decode emotions in the same manner. For this reason, the current study will investigate the possibility of a decoder by emotion interaction and will not assume that emotions are decoded similarly by individual persons.

Assumptions about Emotions

In addition to being explicit about one's assumptions regarding the nature of decoding patterns as discussed above, it is also vital that researchers are very clear regarding their assumptions about emotions in general. In any research, methodologies are in large part based upon researchers' underlying assumptions (either explicit or implicit) about the nature of their topic. Spackman, Brown, and Otto (2009) warn that this is also the case with research on emotions, but that too many researchers are not explicit—if not entirely unaware—of their assumptions about the nature of emotions. For this reason, it is important to make clear exactly what assumptions have been made by researchers historically and which assumptions are made in the current research.

One area where assumptions have been made, perhaps even unknowingly, by many researchers, is in the various approaches to obtain "clean" representations of emotion expression. For example, many studies use trained speakers in an attempt to record "pure" emotions. Such methods are driven by the assumption that emotions exist as discrete entities (for if an individual emotion did not exist in its own discrete domain there could be no single "clean" or "pure" representation of it).

The purpose of the present study is not necessarily to argue that assuming a discrete model of emotion is inherently wrong. It is, however, important that researchers are careful about what assumptions they make regarding the nature of emotions. As always, there are ramifications to these assumptions of which researchers might not be aware.

This sentiment is similarly shared by Klaus Scherer (although he is ironically one of the researchers to whom Spackman, et al. presumably directed their remarks) who suggests that, when venturing into any study of vocalized emotion, one should first understand the theoretical implications of use of the term "emotion" (see Scherer, 2003). For this reason it is necessary to discuss in this paper some of the different theoretical approaches to emotion. There are so many different ways to talk about what emotions are that this paper could not possibly detail all of them. Nor will it make any attempt to do so. Instead, the reader is referred to the following books for more extensive readings on the nature of emotions than can be covered here: Cornelius, 1996; Frijda, 1986; Oatley, 2004; Solomon, 1993.

What Emotions are Not. There exists no simple answer as to what an emotion "is". Therefore, it seems appropriate to begin this discussion with how emotions can be distinguished from what they are not. Scherer (2000) suggests that one problem with much of the work on emotion expression is that emotions are not always kept distinct from other affective states (see also Frijda, 1993). Scherer suggests there exists a five-level hierarchy of affective states with emotions at the top then moving downward to moods, interpersonal stances, attitudes, and personality traits. Scherer suggests there are seven elements on which these constructs differ, as shown in Table 2. Of these seven elements, research has traditionally focused on two, duration and intensity. This research has shown that emotions typically do not endure over long periods of time and are relatively intense. On the other end of the spectrum, affective personality traits are generally quite stable over time and of relatively low intensity.

Table 2
Definitions of different affective states (adapted from Scherer et al., 2001)

Type of affective state: brief definition (*example*)	Intensity	Duration	Synchro-nization	Event focus	Appraisal elicitation	Rapidity of change	Behav-ioral impact
***Emotion**:* relatively brief episode of synchronized response of all or most organismic subsystems in response to the evaluation of an external or internal event as being of major significance (*angry, sad, joyful, fearful, ashamed, proud, elated, desperate*)	**2 - 3**	**1**	**3**	**3**	**3**	**3**	**3**
***Mood**:* diffuse affect state, most pronounced as change in subjective feeling, of low intensity but relatively long duration, often without apparent cause (*cheerful, gloomy, irritable, listless, depressed, buoyant*)	**1 - 2**	**2**	**1**	**1**	**1**	**2**	**1**
***Interpersonal stances**:* affective stance taken toward another person in a specific interaction, coloring the interpersonal exchange in that situation (*distant, cold, warm, supportive, contemptuous*)	**1 - 2**	**1 - 2**	**1**	**2**	**1**	**3**	**2**
***Attitudes**:* relatively enduring, affectively colored beliefs, preferences, and predispositions towards objects or persons (*liking, loving, hating, valuing, desiring*)	**0 - 1**	**2 - 3**	**0**	**0**	**1**	**0 - 1**	**1**
***Personality traits**:* emotionally laden, stable personality dispositions and behavior tendencies, typical for a person (*nervous, anxious, reckless, morose, hostile, envious, jealous*)	**0 - 1**	**3**	**0**	**0**	**0**	**0**	**1**

0: low, 1: medium, 2: high, 3: very high, -: indicates a range

The other five dimensions identified by Scherer (2000) that differentiate emotions from other affective states are the degree to which bodily systems coordinate with each other, the degree to which a state is triggered by some particular event, the degree to which appraisal played a role in eliciting a state, the rapidity of change in the nature of the state, and the degree to which the state affects behavior. Based on the differentiations in these seven dimensions, Scherer suggests research on emotions should focus on those affective states that are relatively intense, short-lived, event specific, etc. Otherwise, one would not be studying emotions at all.

Defining Emotions. With a basic understanding of what emotions are not, one can begin to extrapolate upon what emotions are. Emotion researchers have historically adopted one of two approaches to defining emotion: discrete or dimensional. More recently, a third conceptualization of emotion, called a componential approach, has received support in the literature. The current study adopts such a view of emotion.

Each of these three conceptualizations and a few of the ramifications of their assumptions will be discussed below. These discussions are especially relevant to the present study because the majority of past research has been driven by atheoretical predictions for vocal patterning and the present study attempts to avoid this weakness. It is likely that this atheoretical approach is at least in part responsible for the mixed findings reported in the emotion expression literature, as described above. Discussion of these issues will also help to clarify why four emotions, anger, fear, happiness, and sadness, were selected for investigation in the present research.

Discrete Emotion Theories. Most discrete emotion theories promote the existence of “basic” or “fundamental” emotions (see Darwin, 1872; Tomkins, 1962, Ekman, 1992; Izard, 1971). That is, these theories suggest that a small subset of primitive emotions exists and that these primitive emotions can be combined to create all other emotions. Discrete emotion theorists suggest there are a limited number of basic emotions, the number of which can vary drastically between theories (anywhere from two to fourteen). Ortony and Turner (1990) compared the basic emotions defined in fourteen of the most commonly referenced discrete emotion theories. A summary of their findings can be found in Table 3 (see references to studies in the table). The table suggests that the six most frequently cited basic emotions are fear, anger, disgust, joy, sadness, and surprise.

Table 3

Comparison of basic emotions from several discrete emotion theories (adapted from Ortony & Turner, 1990)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	# of studies
fear	x	x			x	x	x			x	x	x	x		9
anger	x	x			x		x		x		x	x			7
disgust		x			x		x		x		x	x			6
joy		x		x	x						x	x			5
sadness	x	x							x		x			x	5
surprise		x	x		x						x	x			5
happiness			x						x					x	3
interest			x		x							x			3
love	x					x							x		3
rage						x				x			x		3
anxiety				x					x						2
contempt					x							x			2
desire	x		x												2
distress					x							x			2
shame					x							x			2
wonder			x				x								2
acceptance											x				1
anticipation											x				1
aversion	x														1
courage	x														1
dejection	x														1
despair	x														1
elation							x								1
expectancy										x					1
grief						x									1
guilt					x										1
hate	x														1
hope	x														1
pain								x							1
panic										x					1
pleasure								x							1
rage and terror				x											1
sorrow			x												1
subjection							x								1
tender emotion							x								1

Note: Study 1: Arnold, 1960; Study 2: Ekman, Friesen, & Ellsworth, 1982; Study 3: Frijda, 1986; Study 4: Gray, 1985; Study 5: Izard, 1977; Study 6: James, 1884; Study 7: McDougall, 1926; Study 8: Mowrer, 1960; Study 9: Oatley & Johnson-Laird, 1987; Study 10: Panksepp, 1982; Study 11: Plutchik, 1980; Study 12: Tomkins, 1984; Study 13: Watson, 1930; Study 14: Weiner & Graham, 1984

Discrete emotion theories have experienced the most popularity among emotion researchers throughout most of psychology's history. This has especially been the case in research on facial expression of emotion (see, e.g., Ekman, 1992; Izard, 1971). Although less explicit with their assumptions, this is also the case with vocal expression researchers who typically focus on the effects of happiness, sadness, fear, anger, and surprise (see Scherer, 2003 for a review). The model is advantageous in that limiting the number of fundamental emotions allows researchers to focus the targets of their research on only a few emotions. It is generally easy, for example, to differentiate between the patterns of expression for happiness and those for anger. It is significantly more difficult, however, to differentiate between glee and delight.

Most discrete emotion theorists would likely argue that this is simply because glee and delight are very similar emotions. Therefore, glee and delight would have similar profiles. In addition, this approach to emotion is advantageous because most vocal expression studies involve some type of question surrounding the presence of unique acoustic profiles of emotions. It seems to make sense to identify these profiles in more basic emotions first (see Ekman & Davidson, 1994, for discussion).

A serious problem with discrete emotion theories, however, is that there is very little agreement among researchers on what the most primitive emotions are or should be. More recently the concept of the "big six" emotions, that there exist six fundamental emotions from which all other emotions can be created, has enjoyed some popularity. Unfortunately, even among big six proponents, there is no consensus on what the six emotions should be. In fact, the man most often credited with the big six, Paul Ekman (see Ekman, Sorenson, & Friesen, 1969), has himself expanded his list of basic emotions to include a total of twelve (Ekman, 1999).

This is problematic not only from an ontological perspective. It also causes difficulties with research methodologies. Because basic emotions purportedly exist fundamentally to themselves, that is they are not made up of parts of other emotions, identifying discrete profiles becomes problematic when one researcher suggests anger is a basic emotion and another researcher suggests anger is comprised of some combination of pleasure and pain (e.g., Mowrer, 1960).

Dimensional Emotion Theories. Wilhelm Wundt (1874/1905) is traditionally recognized as the first proponent of a dimensional theory of emotion in modern psychology. Although traditionally less popular among emotion expression researchers, dimensional emotion theories began to receive more attention in the mid- to late-1990's due primarily to the work of Bachorowski and Owren (see Bachorowski & Owren, 1995; Bachorowski, 1999). Dimensional theories suggest that emotions exist on a continuous spectrum and can typically be mapped onto two or three orthogonal dimensions. The most common two-dimensional models include as their axes a valence dimension (the degree to which an experience is pleasant or unpleasant) and an activity dimension (the degree to which an experience is active or passive). One of the more frequently referenced two-dimensional models is Russell's (1980) circumplex model (see Figure 1). In three-dimensional models the third dimension is usually something like power or control (see Scherer, 2000 for a review).

In a sense, dimensional models have an advantage over discrete models in that they are less concerned with the existence of distinct borders between emotions. Instead, emotions exist on an affective continuum in which a person can feel more happiness or more sadness but the area between the two emotions is gray. However, dimensional models do not lend themselves well to analyses of emotion expression (at least not to the three questions most commonly discussed in the literature nor the question addressed in the current research) because the "gray area" around emotions disallows the identification of differential patterning.

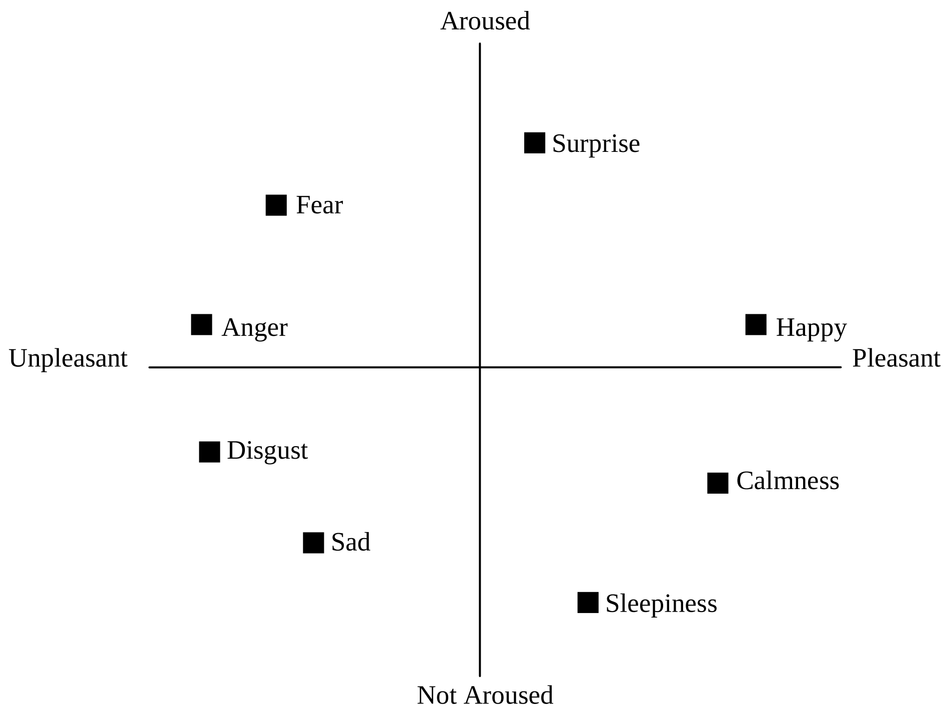


Figure 1: The Circumplex Model (simplified from Russell, 1980)

Componential Emotion Theories. More recently, a third theoretical approach designed to more readily link emotion elicitations to response patterns, called the componential approach, has become popular among many theorists. Componential emotion theories are similar to dimensional theories in that they typically incorporate multiple dimensions (e.g., valence, arousal, control, etc.) into their explanations of emotions. They are also similar to many discrete theories in that, although they do not espouse the theory of basic or primitive emotions, most componential theories assert that there do exist some overarching emotion "families".

Componential theories may be distinguished from the other two theories discussed above not only in the sense that they comprise elements of multiple theories, but they also incorporate another very important aspect. One assumption common to all componential theories is that emotions are the product of cognitive appraisals (though, depending on the model, these need not be conscious or controllable appraisals) of one's environment. It is these appraisals specifically that differentiate the dimensions described in componential models from those in dimensional models. Componential models typically define their dimensions based on different types of cognitive appraisals.

Not all componential models propose the same number of appraisal dimensions, Roseman's (1984) model incorporates several of the most frequently identified dimensions. Roseman suggests that a total of 14 emotion prototypes can be produced from a set of five appraisal dimensions. He names these dimensions 1) Situational state (the degree to which the events a person encounters in a particular situation are consistent or inconsistent with her or his motives in that situation), 2) Probability (how certain or uncertain a person is that something will occur), 3) Agency (the issue of who should be held responsible for an event), 4) Motivational state (whether a person would likely obtain some kind of reward or punishment for an event), and 5) Power (a person's perception of being weak or strong in a given situation). A diagram of Roseman's theory is presented in Table 4.

As mentioned above, Roseman's theory suggests there are 14 primary emotion types. Similar to the discrete emotion theories discussed above, the number of primary emotion families is not agreed upon by componential theorists. Roseman's theory is one of the more moderate theories in the number of emotion families it proposes but other theories can range from only a handful of emotions (e.g., Lazarus, 1991) to virtually an infinite number (e.g., Scherer, 1982; 1993).

Although the number of emotion families or prototypes is not agreed upon by all componential emotion theorists, componential theories avoid the difficulty of discriminating between individual emotions experienced by basic emotion theories because they deal more with groups or categories of emotions. Although one component theory might, for example, identify rage as a fundamental emotion in a particular emotion family and another theory might propose anger in lieu of rage, the two emotions likely still exist within the same family of emotions. Therefore, Scherer (2000) suggests this is perhaps the most appropriate theoretical approach to take when studying vocalized emotion expression.

Table 4

Appraisal dimensions associated with fourteen emotions (from Roseman, 1984)

	Positive		Negative		
	Motive-Consistent		Motive-Inconsistent		
	Appetitive	Aversive	Appetitive	Aversive	
Circumstance-Caused Unknown	Surprise				
Uncertain	Hope		Fear		Weak
Certain	Joy	Relief	Sorrow	Discomfort, Disgust	Weak
Uncertain	Hope		Fear		Strong
Certain	Joy	Relief	Sorrow	Discomfort, Disgust	Strong
Other-Caused Uncertain Certain	Liking		Disliking		Weak
Uncertain Certain			Anger		Strong
Self-Caused Uncertain Certain	Pride		Shame, Guilt		Weak
Uncertain Certain			Regret		Strong

Another primary advantage of utilizing componential emotion theories as a basis for emotion expression research is they lend themselves well to evaluations of the mechanisms that are supposed to underlie the relationship between voice and emotions because they provide empirically testable explanations for the origins of emotion. For a deeper discussion of the advantages of componential models in emotion expression research see Scherer, 2003.

Theoretical Traditions in Emotion Research. Another way to look at emotions is to ask how we as humans experience them. It is one thing to question how emotions should be defined as unembodied entities but it is an entirely different thing to question their existence as part of being human. The latter approach is also important to the current investigation because how a researcher defines emotion in this way determines in some sense what he or she can conclude about the nature of emotions.

This very brief discussion of a few of the major traditions in emotion research is designed neither to be comprehensive nor is it to be particularly critical in its analysis of the various theories. Instead, these theories are mentioned because it is important to recognize what exactly one is studying when one studies expressions of emotion. For example, as will be discussed shortly, some theorists suggest emotions exist most fundamentally as expressions, others suggest they are cognitions, and others suggest they are mere artifacts of our culture. It should be clear, then, how conclusions about the nature of emotions based on emotion expression research under each of these various theoretical approaches can differ significantly.

Darwinian Emotions. Charles Darwin recounts that when surrounded by a fierce mob, Louis XVI famously quipped, “Am I afraid? Feel my pulse.” (Darwin, 1872, p. 238). Darwin used this quote to illustrate his theory that few individuals can mask the expressions of their feelings. More fundamentally, Darwin (and neo-Darwinian theorists) would argue that emotions actually exist at the level of expression (see, e.g., Darwin, 1872; Ekman, 1972; Izard, 1971).

Jamesian Emotions. Similarly, William James (1884) is famous for suggesting that, when evading a bear in the woods, we fear because we run; we do not run because we fear. In other words James, similar to Darwin, suggests that emotions are not felt until they are somehow first expressed. For James (and neo-Jamesian theorists), however, the emphasis is more on the feelings associated with bodily changes than with the actual expression of emotion (see, e.g., Lange & James, 1922; Hohmann, 1966; Schachter & Singer, 1962).

Emotions as Cognitions. Not all theorists, however, believe that emotions exist at such a tangible level. In fact, the cognitivist perspective of emotion, which was perhaps the most dominant theory of emotion in the latter half of the Twentieth Century and, to a certain extent, today, suggests that expressions have very little to do with what emotions actually are. Instead, cognitivists argue, expressions are just what their name implies—outward manifestations of inner cognitive functioning. During the cognitive revolution of the 1960's, Magda Arnold (1960) proposed a cognitive theory of emotions in which she suggested emotions were the result of situation appraisals. Although Arnold was certainly not the first cognitive emotion theorist (see Solomon, 1993, and Lazarus, 1991, for historical accounts of cognitive emotion theory including discussions of Aristotle and Epictetus), she is credited with being the primary pioneer of modern cognitive emotion theory. Although most cognitive theories of emotion differ, often significantly, from Arnold's, they all have in common the primary role of appraisal (see, e.g., Lazarus 1991; Mandler, 1975; Oatley & Johnson-Laird, 1987; Ortony, Clore & Collins, 1988; Roseman, 1984).

Emotion and the Brain. As technology has improved over recent years, cognitive psychology has evolved into cognitive neuropsychology. Although technology has increased neuropsychologists' ability to study how emotions are experienced in the brain, neuropsychological approaches to emotion are not new. In 1937, James Papez proposed what has come to be known as the Papez Loop (see Papez, 1937; LeDoux, 1986). The loop was important to the neuropsychology of emotion because it identified a possible circuit by which emotional meaning gets added to the analysis of incoming stimulus information. Although modern neuropsychologists now identify specific brain structures, such as the amygdala, as locations where humans experience emotions (see LeDoux, 1987), their general explanation of emotions is the same; they suggest that emotions exist at a neurological level.

Emotions as Social Constructs. A final approach to emotion is the social constructivist approach. In contrast to the other approaches briefly mentioned above, social constructivists argue that emotions exist at the most fundamental level not in the body but rather in society. As James Averill, one of the most prominent social constructivists, would suggest, emotions might incorporate elements of biological systems of behavior, but they are not created there. Instead, he would argue, they are created by society to fit the particular needs of the environment (Averill, 1980).

Summary. As mentioned above, these theories are discussed briefly not to provide exhaustive accounts or to laud one approach over another. Instead, these theories are mentioned because it is important to recognize what exactly one is studying when studying expressions of emotion. From a neo-Darwinian perspective (and perhaps from some neo-Jamesian perspectives) the study of emotion expression is in fact a study of emotions themselves. From the other perspectives mentioned, emotion expressions are merely outward manifestations of some other, more fundamental experience.

This differentiation will be important to understand when interpreting results of the current research. It should be noted that the purpose of the current paper is not to address what emotions "are" but instead to help determine how emotions are experienced by perceivers of emotional expressions.

Decoding of Expressions and the Current Research

Of the many ways to look at the experience of emotion expressions, the current study is concerned with the nature of the decoding process. More specifically, this study is concerned with the types of things perceivers of vocalized emotions focus upon to identify those emotions. As evidenced in the discussion of accuracy above, most research on decoding of vocalized expressions of emotion has focused primarily on the listener's ability to appropriately identify vocalized emotion. Although accuracy is among the issues looked at in the present study, it is secondary to the question of whether there exists some set pattern of acoustic cues to which everyone (or most everyone) is acutely attuned or whether patterns in vocalized emotion identification are more individualized. That is to say, do perceivers decode vocalized emotion more nomothetically or idiographically?

As discussed above, although several studies have addressed the issue of whether there exist vocal profiles of encoded (that is, recorded vocalizations of) emotions, to this author's knowledge, no direct study of patterns of acoustic cues used by decoders to identify vocalized emotions currently exists in the literature. For this reason, the current study is primarily exploratory in nature. However, given recent research indicating that acoustic encoding patterns are idiographic in nature, rather than nomothetic, as has been predominantly assumed by emotion researchers (see Spackman, Brown, & Otto, 2009), hypothesis one holds that perceivers of vocalized emotion will focus upon emotion cues in unique and varied ways. That is, decoders will have unique, idiographic ways of utilizing cues to identify emotions, suggesting that decoders vary in their individual implicit theories of how vocal cues correspond to emotions.

To test this hypothesis, the Brunswikian lens model (Brunswik, 1952/1956) will be employed. This model is described in more detail in the method section below so here it is sufficient to state that the lens model can be utilized to measure the degree to which certain cues mediate the relationship between encoded and decoded emotions. In the current study, these cues are measured acoustic properties. As will be detailed more fully below, these properties group into four acoustic domains: pitch, intensity, rate, and spectral characteristics. The method for deriving acoustic properties from the recordings is likewise discussed in greater detail in the method section below.

If the idiographic hypothesis (hypothesis one) holds, lens model statistics should be considerably lower when calculated from the combined emotion identifications of several decoders than when calculated individually for each decoder. On the other hand, if all decoders are attending to cues in the same manner, this consistency should be reflected in a uniform and coherent pattern, and lens model statistics (which indicate the extent to which acoustic properties account for decoder accuracy) should be about the same whether calculated on combined data, or calculated for each decoder individually. If each decoder has a unique way of identifying emotions from acoustic cues, there should be a significant difference between individual and grouped lens model statistics, with individual lens model statistics indicating a substantially higher percent of accuracy explained than when data are combined. Therefore, hypothesis one holds that individual lens model statistics will indeed be significantly higher (in the statistical sense) than grouped ones, indicating that the decoding process varies from person to person.

In hypothesis two, the question of a decoder by emotion interaction will be addressed. Again, because of the lack of literature in this area, directional hypotheses are difficult to derive from past work. However, it makes sense to suggest that if, as hypothesis one states, decoders differ as to which acoustic properties they focus upon to distinguish among emotions, some persons' methods might be more effective for identifying some emotions and other persons' methods more effective for identifying others. Additionally, recent research suggests that some encoders are better at expressing some emotions than they are at others (see Spackman, Brown, & Otto, 2009). If this is the case, it also makes sense to suggest that some decoders would be better at identifying some emotions than others. Therefore, hypothesis two in this study states that there is a decoder by emotion interaction in the analysis of accuracy scores. That is, some persons are better at decoding some emotions and other persons at decoding other emotions.

Hypothesis three also tests the decoder by emotion interaction. In this analysis, however, lens model statistics will again be utilized to test the degree to which this decoder by emotion interaction is mediated by the four domains of acoustic properties. These domains will be included as dependent variables in a multivariate analysis of variance to determine the degree to which acoustic properties of vocalized emotions act as mediators in this interaction.

Finally, although no specific hypotheses are held, the question of the existence of a decoder gender by emotion interaction is addressed. This gender investigation is entirely exploratory in nature and is of interest to the current research simply because, to this author's knowledge, the question has not been previously addressed in the literature.

Method

Encoding

In the current study, vocalized emotions were encoded by sixteen (eight female, eight male) participants. Half the encoders (four female and four male) were recruited from the Performing Arts Department at Brigham Young University (BYU) (Females M_{age}=23.38, $M_{years\ theatre}$=5.63; Males M_{age}=22.25, $M_{years\ theatre}$=5.25). Although not necessarily trained explicitly in the portrayal of emotion, these participants all had several years of acting training which generally emphasizes effective emotional expression. These "trained" encoders were recommended by their department faculty as students who were particularly good at expressing emotions vocally. The other eight encoders (four females and four males) were "untrained" students recruited from undergraduate psychology courses at BYU (Males M_{age}=19; Females M_{age}=18.5). In order to qualify as being untrained, participants could not be performing arts majors and could not have had any formal theater training.

Encoders were requested to read a script multiple times, each time expressing anger, happiness, sadness, or fear. The words in the script were selected to be affectively neutral yet still amenable to the portrayal of each of the selected emotions. The script read as follows:

> It was the first day of school so I left early to see who would be in my required physical science class. I walked in and sat in the middle row. Pat came in and sat next to me. Professor Smith handed out the syllabus and I knew this semester would not be like any other.

The emotions used in this study were chosen for two reasons. First, these are four of the most commonly studied emotions in psychology today so they have relatively well developed vocal profiles. Second, these emotions are typically referenced as among the more primary or fundamental emotions. Therefore, their acoustic properties should be generally representative of more "peripheral" emotions within their families (see discussion of componential emotion theories above). Also, because there is still little known about differential patterning in emotion identification, it makes sense to focus first on these more common fundamental emotions.

Recordings were made independently by each encoder in an anechoic chamber in the basement of the Eyring Science Center on the BYU campus. Inside the anechoic chamber were two stools, a Shure SM58 microphone on a stand, a Kenwood 8 input mixer and a Dell Latitude laptop computer. Voices were recorded at 44100 Hz with 256 bit sampling.

Once the encoder and research assistant were inside the chamber, a sound check was performed to ensure volumes did not exceed recording thresholds. The research assistants ensured that the encoders maintained a distance of approximately two inches from the microphone for recordings. Encoders and research assistants sat in opposing corners of the chamber for the recordings. A Latin square design was employed to vary the order of encoding for each participant.

Each encoder read the script a total of twelve times, three times for each of the four emotions, summing to 192 total recordings over all 16 encoders. Upon completing all twelve recordings, each encoder was requested to rate each of his or her recordings for each emotion as her or his “best”, “middle”, or “worst” expression of the emotion. For this study, recordings self-rated by encoders as being their “middle” expression were dropped from further analysis. This seemed appropriate for three reasons. First, it seemed appropriate to keep the worst portrayals (rather than the middle portrayals) because it was assumed that within emoter variance would be greatest when using the best and worst portrayals. More within emoter variance was preferred because having a higher error term would produce more conservative results. Second, dropping a third of the recordings made the decoding task (described below) much more feasible. Third, two recordings of each emotion provided adequate degrees of freedom within the encoder by emotion interaction to run the appropriate statistical analyses.

Quantification of Acoustic Properties

After all 16 encoders completed their recordings, Praat (Boersma & Weenink, 2005), a computer program designed to synthesize acoustic properties of speech, was employed to create scores for each encoder on 17 acoustic properties (see Table 5 for more complete descriptions of these properties). Seven of the properties are in the domain of the rate of the narration. Rate refers to the overall time duration of the recording as well as the amount of time within this total duration that is divided between speech and pause. Another three of the properties are in the domain of intensity, which refers to the amplitude of the sound wave. Three additional properties are in the domain of pitch, or the fundamental frequency (F0) of the vibration of the vocal folds. The final four acoustic properties are measures of spectral characteristics. The label “spectral characteristics” refers to those qualities that affect the perception of voice quality or timbre (including the ratio of high vs. low frequency harmonics, etc.)

Table 5
Acoustic dimensions and definitions of their associated measured properties

Rate		
1.	Total utterance duration	Sum of duration of all speech segments
2.	Average utterance duration	Mean duration of each individual speech segment
3.	Standard deviation utterance duration	The degree to which all individual speech segments deviate from the average utterance duration
4.	Total pause duration	Sum of duration of all utterances
5.	Average pause duration	Mean duration of each individual pause segment
6.	Standard deviation pause duration	The degree to which all individual pause segments deviate from the average pause duration
7.	Total narration duration	Sum of total utterance duration and total pause duration
Intensity		
8.	Average intensity	Mean intensity of total narration duration
9.	Standard deviation intensity	The degree to which intensity varies about the average intensity
10.	Range of intensity	The span between the lowest and highest intensity ratings
Pitch		
11.	Average pitch	Mean F0 of total narration duration
12.	Standard deviation pitch	The degree to which F0 varies about the average F0
13.	Range of pitch	The span between the lowest and highest F0 ratings
Spectral Characteristics		
14.	Center of gravity	Mean wave frequency across the power spectrum over entire emotion recordings
15.	Spectral standard deviation	The degree to which frequencies in a spectrum deviate from the center of gravity
16.	Spectral skewness	Measure of how different the shape of the spectrum below the center of gravity is from the shape of the spectrum above the center of gravity
17.	Spectral kurtosis	Measure of the degree to which the shape of the spectrum around the center of gravity deviates from the normal distribution

The four dimensions of rate, intensity, pitch, and spectral characteristics were chosen for study in the current research because of their frequent use in previous studies of the acoustic properties of vocalized emotion (e.g., Banse & Scherer, 1996; Juslin & Laukka, 2003; Murray & Arnott, 1993; Scherer, 2003; Scherer et al., 1991). In the current study the intensity threshold for differentiating between pause and speech was set at 35dB. A sampling rate of 10 milliseconds was used for the analyses of all acoustic properties.

Decoding

All 128 recordings were saved as mp3 files and embedded into a Microsoft Access program to be listened to and identified by a second group of participants (referred to as decoders throughout the remainder of this paper). Eighteen college aged decoders, ten female and eight male, were recruited from undergraduate psychology courses at BYU. Each decoder signed up for one of several possible decoding sessions held in a large computer lab. Computers in the lab were equipped with a RealTek 260 sound card. As decoders arrived at the lab, research assistants loaded the appropriate sound files onto each decoder's computer. Once all decoders for a particular session arrived, the research assistant provided instructions on how to proceed through the Access response application. After providing the instructions, the research assistants distributed Sony MDR-101LP headphones so that decoders could listen to each recording at their own pace without bothering their neighbors.

The order of recordings was randomized in the Access program. Each decoder listened to all 128 recordings and identified which emotion was expressed in each. Decoders were provided with five options: anger, fear, happiness, sadness, or I don't know. Upon identifying the emotion, decoders indicated the degree to which they were certain they had identified the appropriate emotion by checking a box along a nine-point scale ranging from completely certain to not at all certain (see Appendix A for a sample screen shot).

Prior to listening to all 128 recordings, decoders listened to a sample voice recording of the script to get a sense of what they should expect. Decoders could replay each emotion as many times as they wanted before making their determination and moving on to the next recording. Most decoders completed the task in 45 minutes to an hour. After listening to all the recordings, decoders filled out a demographics page. Extra credit was provided for decoders who submitted names of professors offering extra credit for research participation. Otherwise, no compensation was provided.

After several decoders had completed their ratings, it was found that the order of encoded emotions had not been randomized appropriately. Data from these decoders were not used in the current analysis[2]. Therefore, all analyses were performed on data from eight (four male, four female) decoders. Although this was not a large sample, given the idiographic nature of the hypotheses evaluated, the exploratory nature of this study, and the large number of observations for each decoder, the sample size seemed adequate. In addition, because of the large number of portrayals (128) identified by each decoder, this sample provides ample degrees of freedom in all of the analyses.

Analysis

Grouped vs. Idiographic Lens Model Statistics. The first hypothesis, regarding the idiographic nature of emotion decoding, is tested using the Brunswikian lens model. First developed by Egon Brunswik (1952; 1956), the lens model enables the researcher to identify the cues that lead to a person's perceptions of some event. The model assumes that persons have at their disposal any number of cues from distal variables that could lead them to make a given perceptual judgment. The cues measured in the present research are the seventeen acoustic properties of the encoded emotions (the distal variables). Therefore, in this study the lens model was used to quantify consistency in the acoustic mediation of emotion judgments. In other words, results from the lens model identified the extent to which particular acoustic properties

[2] It should be noted that the author intends to utilize these data in future research if it can be deemed they are useful.

are used by decoders to correctly identify the encoded emotions. It should be noted that the lens model is particularly appropriate for the present study because it was designed specifically for identifying idiographic judgment patterns within individuals (see, Beal et al., 1978; Bernieri, Gillis, Davis, & Grahe, 1996; Brunswik, 1952, 1956; Cooksey, 1996).

Figure 2 presents a diagrammatic representation of the lens model adapted from Spackman, Brown, & Otto (2009). This representation shows the relationship between the proximal judgment or perception (decoded emotion), the cues (acoustic properties), and the distal variable (encoded emotion). Algebraically, the model can be represented as follows: $r_a = \mathrm{GR}_e\mathrm{R}_s + \mathrm{C}\sqrt{1-\mathrm{R}_e^2}\,\mathrm{C}\sqrt{1-\mathrm{R}_s^2}$. In this equation r_a represents the correlation between the distal variable (Y_s) and the judgment (Y_e), $\mathrm{GR}_e\mathrm{R}_s$ represents the degree to which the cues mediate the judgment, and $\mathrm{C}\sqrt{1-\mathrm{R}_e^2}\,\mathrm{C}\sqrt{1-\mathrm{R}_s^2}$ is the error or the degree to which the cues do not mediate the judgment.

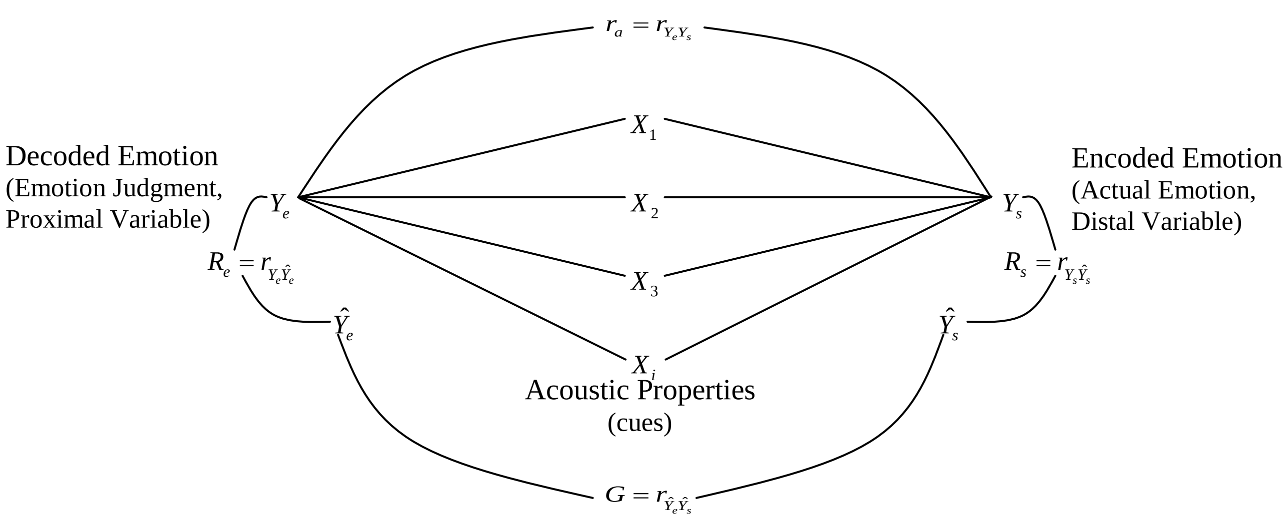


Figure 2: The Brunswikian lens model (adapted from Spackman, Brown, & Otto, 2009)

A primary task of the present study was to determine the degree to which judgments can be accounted for by the acoustic variables. Therefore, the ratio of $\frac{GR_eR_s}{r_a}$ is used to measure the degree to which the acoustic variables can account for judgments. This ratio will be referred to as the identification ratio (see Spackman, Brown, & Otto, 2009, for the initial use of this term). Hypothesis one holds that this identification ratio is significantly higher for individual decoders than it is when calculated on the combined data of all decoders as a group. This hypothesis is tested with a simple two-way analysis of variance, with the comparison of individual identification ratios against the grouped ratios as one factor and acoustic domain as the other factor. If this hypothesis were borne out, it would constitute evidence that decoders make their identifications idiographically. If all decoders make their judgments on the basis of the same acoustic patterns, identification ratio scores would not differ between individual analyses and grouped analyses.

To accommodate the lens model, proximal emotion judgments and distal actual emotion identities are considered as dichotomous pairs (e.g., for anger, 1=”Anger”, 0=”Not Anger”). Because of the binary nature of these variables the correlation coefficients (r_a) are calculated using the phi coefficient. Predictions of encoded and decoded emotions from the acoustic properties therefore utilize logistic regressions. Cooksey’s (1996) lens model equations for dichotomous data, with Stewart’s (2004) correction, are used to calculate lens model statistics from the logistic regression analyses.

Tests of Accuracy. In addition to the identification ratios discussed above, the *d’* statistic is employed to measure decoders’ accuracy in identifying encoded emotions. This statistic is based upon signal detection theory (see Swets, Tanner, & Birdsall, 1961). As can be seen in Table 6, there are four potential outcomes associated with signal and noise in a typical identification task: hit, miss, false alarm, and correct rejection.

Table 6
Matrix of response options for emotion identification

		Response (decoded emotion)	
		Yes	**No**
Stimulus (encoded emotion)	**Signal**	I. Hit (appropriately identified emotion of interest)	II. Miss (improperly identified stimulus as some other emotion)
	No Signal	III. False Alarm (falsely identified encoded emotion as emotion of interest)	IV. Correct Rejection (appropriately identified stimulus as not the emotion of interest)

In the present study, a method for calculating *d'* suggested by Wickens (2001) is used. Recall that, in addition to identifying the emotion, decoders in the present study were requested to indicate the degree to which they were certain their identifications were correct. As suggested by Wickens, the identified emotion can be combined with the decoder's degree of certainty to create the matrix represented in Table 7.

Table 7
Emotion identification/certainty matrix for decoder one, best, anger

	YES									NO									
	Certain								Uncertain	Certain								Uncertain	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	Total
Anger	1	3	3	1		2	3	2							1				16
Not Anger				1			1	1		1	4	8	7	6	7	5	6	1	48

Table 7 represents decoder five's actual responses to the 64 "best" recordings. As was the case for calculating the lens model statistics, when calculating *d'*, encoded (stimulus) emotions are identified dichotomously as either "anger" or "not anger" (or "fear" / "not fear", etc.). The decoded (response) emotions are also identified dichotomously as either "yes" (for anger) or "no" (for not anger). To estimate[3] *d'* statistics from this matrix, scores from each quadrant are summed. Then, summed scores for each quadrant are divided by their respective row totals. These newly created percentages are then standardized and the difference between scores for "hits" and "false alarms" is taken. Therefore, *d'* statistics are interpreted like *z*-scores.

The signal to noise ratio is represented graphically in Figure 3. As can be seen, *d'* is literally an estimate of the distance between the mean of the signal distribution and the mean of the noise distribution. The larger the *d'* distance, the better the discrimination between signal and noise.

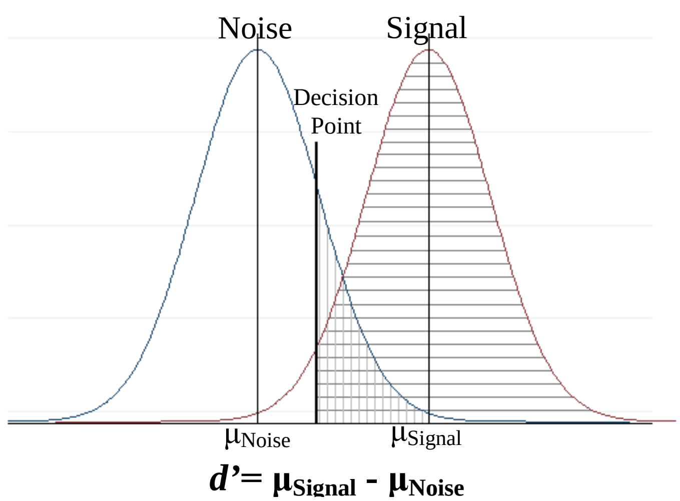


Figure 3: Graphical representation of the *d'* statistic in Signal Detection Theory.
Note: The vertically shaded portion (the area within the noise distribution above the decision point and below the signal distribution mean) represents false alarms whereas the horizontally shaded portion (the area between the signal distribution mean and the decision point) represents hits.

Decoder by Emotion Interaction. To test the second hypothesis, that there exists a decoder by emotion interaction for accuracy, the *d'* statistic for each emotion identification is included as the dependent variable in a two-way analysis of variance (ANOVA) model. The

[3] It should be noted that this method only approximates *d'* statistics. Wickens (2001) describes a method to calculate the exact values by repeating this process with varying decision points to make all possible comparisons of hits to false alarms, then plotting all possible probabilities on a receiver operating characteristic (ROC) curve. This level of accuracy, however, was not deemed necessary in the current study.

independent variables, of course, are decoder ID and encoded emotion. Hypothesis two does not involve acoustic mediation. It is a test of the decoder by emotion interaction using a pure and simple measure of accuracy, the *d'* statistic. Hypothesis three, on the other hand, tests the decoder by emotion interaction using the lens model identification ratio. Therefore, like hypothesis one, hypothesis three is concerned with the acoustic mediation of judgments.

Decoder by Emotion Interaction in Lens Model Statistics. To test hypothesis three, identification ratio data are analyzed with a two-way split plot repeated measures multivariate analysis of variance (MANOVA), with one between subjects factor (decoder) and one within subjects factor (emotion). The dependent variables include the identification ratios for the four acoustic domains (rate, intensity, pitch, and spectral characteristics). In addition, two measures of identification accuracy, the *d'* statistics calculated for hypothesis two and corresponding phi coefficients, will be included as dependent variables. Including these measures of accuracy in addition to the identification ratios allows for a better understanding of the multivariate space in which the decoding process occurs. It should also be noted that, although the two accuracy measures are somewhat redundant of each other, both of these measures are included (as opposed to just one) as dependent variables simply as a cross check to ensure that accuracy rates are being measured appropriately[4].

For each decoder 32 lens models and therefore 32 identification ratios are calculated, eight for each of the four acoustic domains. Each of these identification ratios is comprised of two recordings (one rated as the encoder's best and one rated as his or her worst) of each of the four emotions anger, fear, happiness, and sadness. With eight decoders and 32 identification

[4] In the current study, *d'* and phi coefficients are correlated with each other at $r = .858$, indicating they are indeed both measuring accuracy rates similarly.

ratios per decoder, the MANOVA described above will include 256 identification ratios as the primary data.

Each of these lens model identification ratios is calculated idiographically, that is, on individual decoders. Spackman, et al. (2009) found that most of the identification ratios calculated on encoders idiographically were above .98. For this reason, it was initially anticipated that this might also be the case for decoders. Were this to be the case with decoders also, this kind of ceiling effect could have made the test of hypothesis three described above ineffectual. Therefore, had this been the case, a secondary strategy would have been employed. This strategy would have been to do a similar split plot MANOVA analysis, but with lens model statistics calculated on each of the four acoustic domains separately. However, as will be seen in the results section below, identification ratios for decoders were not large enough to cause this ceiling effect. Therefore, this secondary strategy was deemed unnecessary.

Gender Interactions. Both the two-way ANOVA model of hypothesis two and the MANOVA model for hypothesis three were run with gender included as an additional independent variable to test for potential gender by emotion interactions. However, results of these analyses indicated that gender does not play a significant role in how persons identify vocalized emotions. Because these hypotheses were not central to the study and because results for these analyses were not significant, the gender analyses will not be discussed further in this paper.

Results

Hypothesis One

Spackman, Brown, and Otto (2009) recently demonstrated there is much more consistency in the pattern of acoustic properties of encoded vocalized emotion within an individual encoder than there is within pooled data of many encoders. Therefore, it was hypothesized in the current study that a similar consistency advantage for individual emotion decoding over grouped emotion decoding would exist. This hypothesis was tested with a two-way ANOVA in which identification ratios derived from the lens model were included as the dependent variable. Individual and group identification ratios were dummy coded and included as one independent variable. The four acoustic domains were the other independent variable. Results from the ANOVA indicate there are no significant differences between individual and group identification ratios (F=.022, p=.882, partial η^2=.000). Instead, the ratios are very similar in magnitude, nothing like Spackman, et al.'s (2009) results on the encoded data. This suggests that, although encoders tend to express emotions acoustically in many ways, decoders' methods for interpreting vocal emotion expressions tend to be highly consistent across individuals. In other words, emotions tend to be encoded idiographically but decoded nomothetically.

Although no main effect for individual vs. group identification ratios exists, there is a significant main effect for acoustic domain (F=3.61, p=.015, partial η^2=.074). Results from Tukey's b post hoc analysis[5] on acoustic domain indicate that the mean identification ratio value for F0 (M=.089) is significantly lower than means on the other three domains (Intensity M=.162, Rate M=.212, Spectral Characteristics M=.168). This indicates that, in general, decoders focus less on measures of fundamental frequency (interpreted by the ear as elements of pitch and intonation) to differentiate between emotions than they do on the other acoustic properties.

Results indicate no significant individual vs. group by acoustic domain interaction effect. See Appendix B for all individual and group identification ratios and Appendix C for the complete ANOVA summary table for this analysis.

Hypothesis Two

[5] The reader is referred to Howell, 2002 pp. 391-404 for a discussion of several post hoc and a priori analyses, their underlying assumptions, and suggestions for which analyses are most appropriate for which types of data.

Because it was initially believed that decoders vary in the methods they employ to identify different emotions, it was hypothesized that some decoders would be good at identifying some emotions while other decoders would be better at identifying other emotions. To test this hypothesis, *d'* accuracy statistics were submitted to a two-way ANOVA with decoder and emotion included as independent variables.

Results from this analysis indicate that no evidence for such an interaction between decoder and emotion exists (F=0.88, p=.601, partial η^2=.332). This result should be interpreted with some degree of caution, however. The partial η^2 value is peculiarly high given the low degree of significance. Results from a power analysis (δ=.464) suggest that the non significant F ratio might be due to inadequate sample size.

However, there does exist a significant main effect for decoder (F=3.02, p=.019, partial η^2=.362), indicating that some persons are better at identifying emotions in general than are others. In addition, there also exists a large and significant main effect for emotion (F=11.21, p<.001, partial η^2=.512). Mean *d'* statistics for each decoder and emotion are provided in Table 8. All individual *d'* statistics are provided in Appendix D.

Table 8

Mean *d'* values for decoders and emotions

	Anger	**Fear**	**Happy**	**Sad**	Total
Decoder 1	2.83	1.45	1.61	1.96	1.88
Decoder 2	2.12	1.9	2.41	1.72	2.21
Decoder 3	2.68	1.73	2.44	1.56	2.32
Decoder 4	1.87	1.04	2.08	1.44	1.77
Decoder 5	2.74	1.32	3.65	1.60	2.84
Decoder 6	2.21	1.64	2.4	1.96	2.16
Decoder 7	2.21	1.62	2.36	1.39	2.14
Decoder 8	1.37	1.16	1.9	0.91	1.58
Total	2.25	1.48	2.36	1.57	

Decoder Main Effect. Tukey's HSD post hoc analysis was utilized to determine more precisely how decoders differed in their levels of accuracy. Results indicate that the only two decoders who differed from each other significantly were decoders five and eight (see Appendix E for complete results from the Tukey's HSD post hoc analysis). This suggests that decoders' accuracy levels are distributed fairly evenly. In other words, there does not exist a distinct delineation between "good" and "bad" decoders.

This distribution can more easily be understood by observing how the means provided in Table 8 look graphically. These means are plotted in Figure 4. Examination of this graph suggests that much of the variance between decoders is created by a very high mean for decoder five on happiness and relatively low scores for decoder eight on anger and sadness relative to the other decoders. Otherwise, decoders tend to be fairly similar in their identification accuracy levels.

Emotion Main Effect. It is also clear from Table 8 and Figure 4 that, on average, decoders are better at identifying anger and happiness (Anger $M=2.25$, Happiness $M=2.36$) than they are at identifying fear and sadness (Fear $M=1.48$, Sadness $M=1.57$). These observations are supported statistically with results from Tukey's b post hoc analysis. This analysis compares means for each variable (in this case the four emotions) and groups means of similar value into subsets. In this case, anger and happiness group on subset one and fear and sadness group on subset two, indicating the two groups of emotions differ significantly from each other.

Hypothesis Three

Because no significant decoder by emotion interaction effect was found in hypothesis two it could be argued that hypothesis three (which investigates the degree to which this interaction effect is be mediated by acoustic properties) no longer has applicability. In a sense, this is an appropriate argument. However, there is more to be learned from the proposed analyses than was contained in the original hypothesis three. As will be seen below, perhaps the most important findings from the current study are derived from elements of the MANOVA analysis not directly associated with the decoder by emotion interaction.

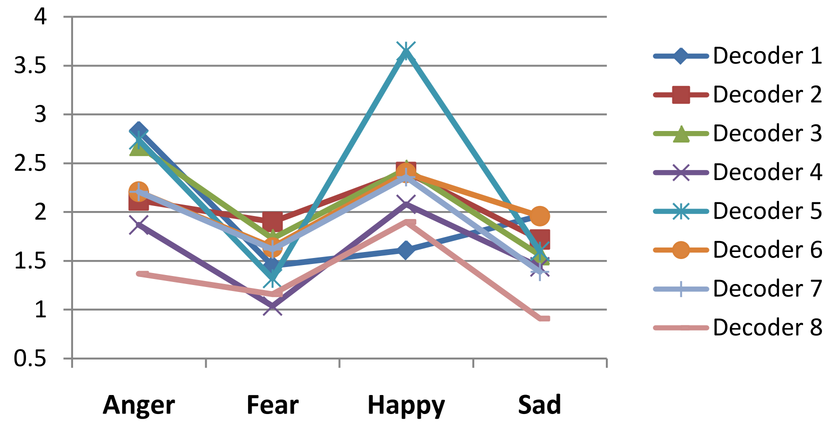


Figure 4: Graph of mean *d'* statistics for each decoder by emotion. *Note* the relatively high score for decoder five on happiness and the relatively low scores for decoder eight on anger and sadness.

The MANOVA model used in this analysis includes three independent variables (decoder, emotion, and gender) and six dependent variables (the four domains of acoustic properties and two measures of accuracy). To create the necessary within decoder variance for the MANOVA model, identification ratios were calculated separately for the "best" and "worst" recordings (see Appendix F for a list of all identification ratios). Due to idiosyncrasies of the Stewart (2004) correction of the lens model, 12 of the 128 identification ratios calculated on the "best" recordings were very close to zero, but actually appeared to be negative. These values were changed to zero in subsequent analyses because it should be mathematically impossible to have a negative identification ratio.

Also, it should be noted that, due to the specific combination of fixed (gender and emotion) and random (decoder) variables, the within error term was not appropriate for all comparisons. Table 9 presents the expected mean squares table for the MANOVA model employed in this study.

Table 9
Expected Mean Squares for MANOVA model

Expected Mean Squared Table			
	Df	EMS	Error Term
Between Person	**7**		
Gender	1	$\sigma_e^2 + n\mathrm{d(g)E}\sigma_{\mathrm{G}}^2$	$MS_{d(g)}$
Decoder(Gender)	6	$\sigma_e^2 + n\mathrm{E}\sigma_{\mathrm{d(g)}}^2$	MS_{Within}
Within Person	**56**		
Emotion	3	$\sigma_e^2 + n\sigma_{d(g)\,e}^2 + n\mathrm{Gd(g)}\sigma_{\mathrm{E}}^2$	$MS_{E\ x\ d(g)}$
Emotion x Gender	3	$\sigma_e^2 + n\sigma_{d(g)\,e}^2 + n\mathrm{d(g)}\sigma_{\mathrm{G\ E}}^2$	$MS_{E\ x\ d(g)}$
Emotion x Decoder(Gen)	18	$\sigma_e^2 + n\sigma_{d(g)\,e}^2$	MS_{Within}
Error Within	32	σ_e^2	
Total	**63**		

Multivariate Results. Results from the multivariate tests indicate three significant effects and are summarized in Table 10. First, there exists a significant main effect for decoder. It is important to note that this effect is only significant on Roy's Greatest Root (RGR). As will be discussed further below, it makes sense that this value would be significant on RGR and none of the other measures because RGR is unidimensional. That is, it is only concerned with the dimension that accounts for the greatest amount of variance explained. In the case of the decoder main effect, most of the variance is accounted for by the accuracy dimension and very little is accounted for by the other dimensions in the dependent variable space. Therefore, it makes sense that decoder would only be significant on RGR.

Second, as shown in Table 10, there exists a significant main effect for emotion on all four multivariate measures. This effect will be discussed in great detail below. Third, there exists a significant decoder by emotion interaction on Roy's Greatest Root. Again, as will be evidenced below, this difference is primarily accounted for by the unidimensional nature of the differences.

Table 10
MANOVA Summary Tables

Statistic	Value	*p*
Decoder		
Wilks' Lambda	0.272	0.2679
Pillai's Trace	1.000	0.3775
Hotelling-Lawley Trace	1.786	0.1908
Roy's Greatest Root	**1.194**	**0.0002**
Emotion		
Wilks' Lambda	**0.001**	**<.0001**
Pillai's Trace	**2.617**	**<.0001**
Hotelling-Lawley Trace	**42.099**	**<.0001**
Roy's Greatest Root	**30.331**	**<.0001**
Decoder by Emotion		
Wilks' Lambda	0.096	0.9399
Pillai's Trace	1.758	0.96
Hotelling-Lawley Trace	3.298	0.8937
Roy's Greatest Root	**1.471**	**0.0086**

Univariate Results. To gain a better understanding of how the independent variables behave on each dependent variable individually, scores on each dependent variable were simultaneously submitted to independent univariate ANOVAs. Rencher and Scott (1990) argue that MANOVA multivariate tests are a good filter to protect against alpha inflation. In this approach, univariate tests are never considered to be significant unless the corresponding multivariate effect is also significant. The authors argue that the multivariate "filter" is effective if any of the four multivariate tests are found to be significant. Therefore, it was deemed appropriate to conduct univariate analyses on those effects that were significant multivariately with some assurance that these tests are protected against alpha inflation. Summary tables of results from these ANOVAs are detailed in Appendix G.

Not surprisingly, (because they are exact replications of the analyses in hypothesis two) there were no significant decoder by emotion interactions on either of the two accuracy measurements (*d'*: F=0.88, p=.601; phi: F=1.17, p=.342). As was also the case for hypothesis two, there were significant main effects for both decoder (*d'*: F=3.02, p=.019; phi: F=5.81, p<.001) and emotion (*d'*: F=11.21, p<.001; phi: F=15.66, p<.001) on scores of accuracy.

Perhaps more importantly, however, results on the four acoustic domains indicate there are no significant decoder by emotion interactions on any of the four acoustic domains univariately (F0: F=0.33 p=.993; Intensity: F=0.39 p=.981; Rate F=0.17 p=1.0; Spectral: F=0.57 p=.895). However, a significant main effect for emotion does exist on three of the four domains (Intensity: F=13.47, p<.001, partial η^2=.558; Rate: F=5.65, p=.003, partial η^2=.346; Spectral Characteristics: F=18.09, p<.001, partial η^2=.629) and approached significance on the fourth domain (F0: F=2.56, p=.072, partial η^2=.194).

Emotion Main Effect. A comparison of means for identification ratios and accuracy statistics by emotion is provided in Table 11 below. Observation of the means in Table 11 makes very clear that particular emotions are defined more by some acoustic properties than by others. Specifically, F0 is used to identify fear more than any of the other emotions, intensity is used to identify sadness, rate is used to identify happiness and sadness more than the other two emotions, and spectral characteristics are used to identify anger. Table 11 also indicates that anger and happiness are more accurately identified by decoders than are fear and sadness (as was discussed in hypothesis two).

Table 11
Means of identification ratios (for four acoustic domains) and accuracy statistics by emotion

	F0	Intensity	Rate	Spectral	Phi	*d'*
Anger	.0961	.1683	.1706	**.3146**	**.651**	**2.254**
Fear	**.1688**	.0853	.2148	.1891	.479	1.481
Happiness	.0697	.1534	**.3656**	.0598	**.672**	**2.358**
Sadness	.0959	**.3183**	**.3699**	.1531	.517	1.566

These observations are supported by results from linear contrast tests comparing how emotion means differ on the various dependent variables. Results from these contrast tests confirm that identification ratios for fear are significantly higher than they are for other emotions on F0 (F=7.03, p=.012), identification ratios for sadness are significantly higher than they are for the other emotions on intensity (F=34.94, p<.001), identification ratios for happiness and sadness together are significantly higher than for the others on rate (F=16.43, p<.001), identification ratios are higher for anger than they are for the others on spectral characteristics (F=39.79 p<.001), and anger and happiness have higher accuracy identification values than do the others (for *d'*: F=33.15, p<.001; for phi: F=45.41, p<.001).

To get a better idea of how the emotions fit together with acoustic properties and accuracy statistics multivariately, identification ratios and accuracy scores for each emotion were plotted together three dimensionally using Metrika, a software system designed ideally for graphing multivariate structures in a three dimensional space. Figures 5 through 9 summarize the nature of this space by illustrating where the four acoustic property domains and the two accuracy measures lie. The figures show very clearly that each of these domains defines unique portions of the multivariate space. In addition, each of the four emotions is mapped on each vector. These figures illustrate graphically the effects of the contrast tests described above.

Ultimately, these figures illustrate which acoustic properties decoders use to identify vocalized emotions. Specifically, decoders identify anger by focusing on spectral characteristics, fear primarily by focusing on frequency (F0), happiness primarily by focusing on rate, and sadness primarily by both intensity and rate.

Decoder and Decoder by Emotion Effects. As mentioned above, there is also a main effect for decoder multivariately. However, as can be seen in Appendix G, the only two dependent variables on which decoder had a main effect were the two accuracy measures (*d'*: F=3.02, p=.019; phi: F=5.81, p<.001). This indicates that the majority of the variance responsible for the multivariate main effect is accounted for by the accuracy dimension. This helps to explain why decoder was only significant on Roy's Greatest Root. On the acoustic properties dimensions, decoders actually show a great deal of consistency in their ratings of emotions. Figures 10a and 10b plot mean identification ratios for each decoder on each of the four emotions within a three dimensional acoustic domain space. As can be seen in Figures 10a and 10b, male and female decoders show remarkable consistency in the ways they utilize specific acoustic properties to identify emotions. These graphs lend further support to the conclusion from hypothesis one that decoders identify vocalized emotions nomothetically.

Discussion

Most studies of vocal expressions of emotion have focused on one or more of three questions: whether there exist specific and unique patterns in the acoustic properties of individual emotions, whether these patterns are universal across cultures, and the degree to which decoders can accurately identify vocal expressions of emotion. Until now, no studies have explicitly addressed the issue of potential acoustic profiles used in the decoding process. The primary purpose of this dissertation is to investigate how decoders use specific acoustic variables in their identifications of vocalized emotions. It is the first study directly concerned with the cognitive processes of emotion decoding.

Figure 5
Summary of F0 Domain

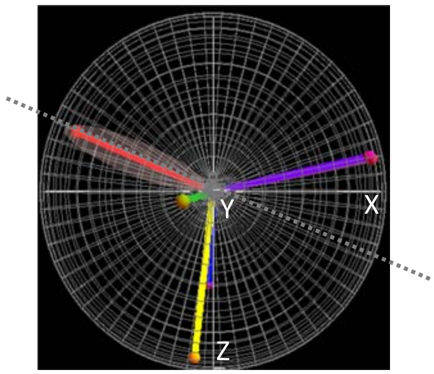


Figure 5a: The F0 (red) vector.
Note: Top view. The X axis is defined by F0 in the negative direction. Some of the z axis is also defined by F0 in a negative direction.

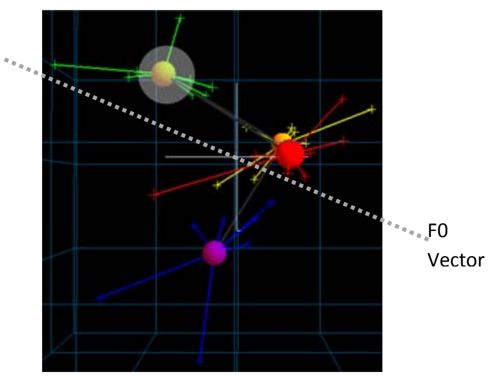


Figure 5b: Emotions on the F0 vector.
Note: Fear is located independently from the other three emotions along the F0 vector.

Emotions	Mean IRs
Anger	9.6%
Fear	**16.9%**
Happiness	7.0%
Sadness	9.6%

Linear Contrast of Fear vs. other 3 emotions:
F=7.03 p=.0124

Figure 6
Summary of Spectral Characteristics Domain

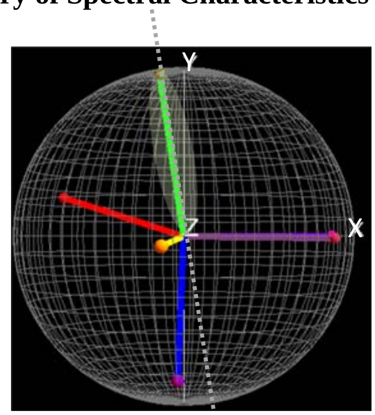


Figure 6a: The Spectral (green) vector.
Note: Front view. The Y axis is defined in the positive direction by spectral characteristics.

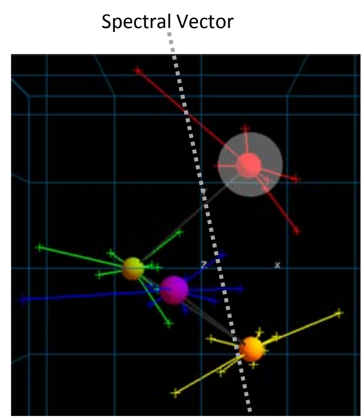


Figure 6b: Emotions on the spectral characteristics vector.
Note: Anger is located independently from the other three emotions along the spectral characteristics vector.

Emotions	Mean IRs
Anger	**31.5%**
Fear	18.9%
Happiness	6.0%
Sadness	15.3%

Linear Contrast of Anger vs. other 3 emotions:
F=39.79 p<.001

Figure 7
Summary of Rate Domain

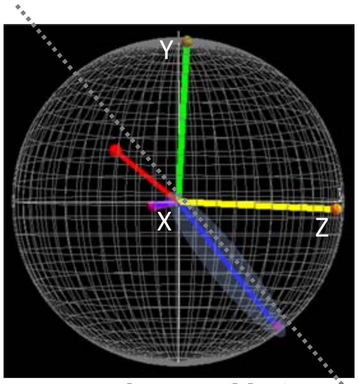


Figure 7a: The Rate (blue) vector.
Note: Side view. The space between the positive Z axis and the negative Y axis is defined by rate properties.

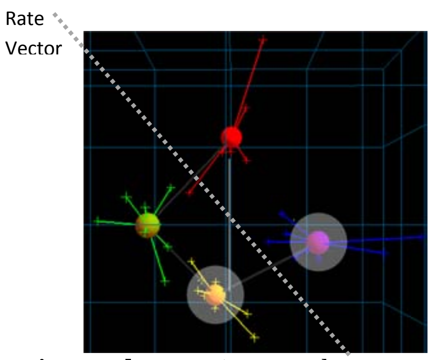


Figure 7b: Emotions on the rate vector.
Note: Happiness and sadness are located independently from the other two emotions along the rate vector.

Emotions	Mean IRs
Anger	17.1%
Fear	21.5%
Happiness	**36.6%**
Sadness	**37.0%**

Linear Contrast of happy & sad vs. other 2 emotions:
F=16.43 p<.001

Figure 8
Summary of Intensity Domain

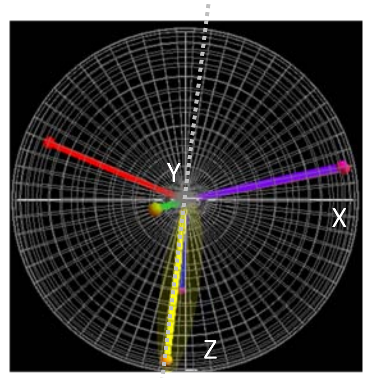


Figure 8a: The Intensity (yellow) vector.
Note: Top view. The Z axis is defined in the positive direction by intensity.

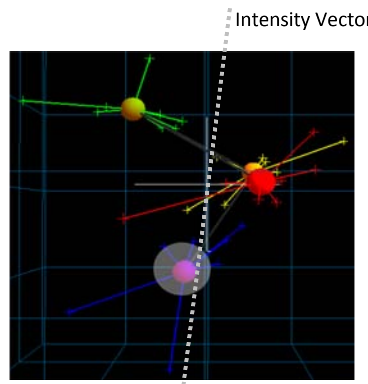


Figure 8b: Emotions on the intensity vector.
Note: Sadness is located independently from the other three emotions along the intensity vector.

Emotions	Mean IRs
Anger	16.8%
Fear	8.5%
Happiness	15.3%
Sadness	**31.8%**

Linear Contrast of sad vs. other 3 emotions:
F=34.94 p<.001

Figure 9
Summary of Accuracy Domain

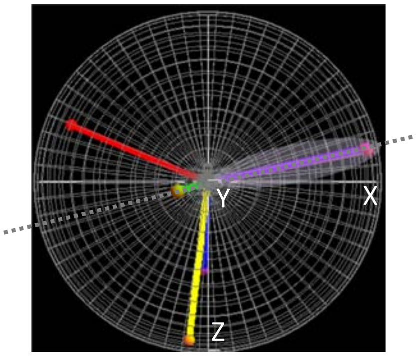


Figure 9a: The Accuracy (purple) vectors.
Note: Top view. There are actually two purple vectors on top of each other (one for *d'* and one for phi). The X axis is defined in the positive direction by measures of accuracy.

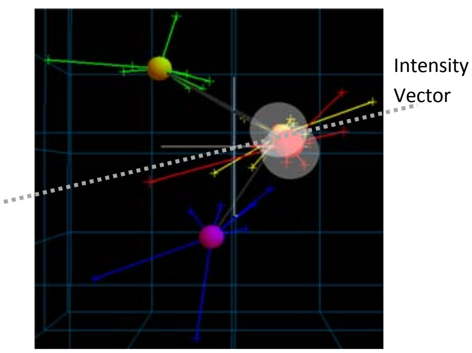


Figure 9b: Emotions on the accuracy vector.
Note: Anger and happiness are located independently from the other two emotions along the accuracy vector.

Emotions	Mean *d'*	Mean Phi
Anger	**2.25**	**.651**
Fear	1.48	.479
Happiness	**2.36**	**.672**
Sadness	1.57	.517

Linear Contrast of anger & happy vs. other 2 emotions:
For *d'*: F=33.15 p<.001
For Phi: F=45.41 p<.001

Figure 10
Plots of Decoder by Emotion

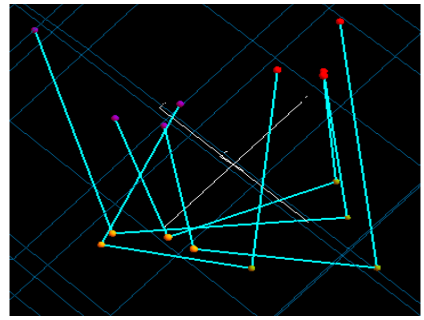


Figure 10a: Plot of male decoders by emotion on the acoustic domains.
Note: The consistent U shapes of these distributions indicate a very high consistency in the way decoders use acoustic properties to identify emotions.

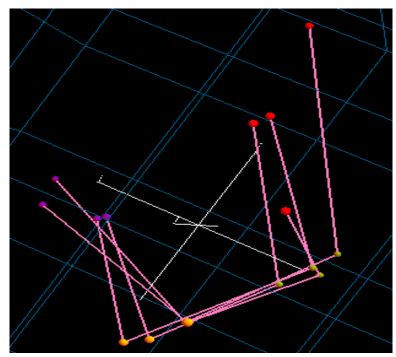


Figure 10b: Plot of female decoders by emotion on the acoustic domains.
Note: The consistent U shapes of these distributions indicates a very high consistency in the way decoders use acoustic properties to identify emotions.

Color key for vector Figures 5a-9a
Green: Spectral Characteristics
Yellow: Intensity
Red: F0
Blue: Rate
Purple (two of them): Accuracy (*d'* and phi)

Color key for emotion stars Figures 5b-9b and 10a-10b
Green: Fear
Yellow: Happiness
Red: Anger
Blue: Sadness

Decoder Accuracy

Decades of evidence indicate not all persons are equal in their ability to accurately identify vocalized emotions (see Scherer, 2003). Recent research provides evidence that some persons are better at encoding some emotions than they are at encoding others (Spackman, Brown, & Otto, 2009). Given these findings, it was hypothesized that a decoder by emotion interaction would also exist, in which individual decoders would be better at identifying some emotions and worse at identifying others. Results indicate that, although decoders vary significantly in their ability to identify vocalized emotions in general, there exists no evidence for a decoder by emotion interaction. Rather, persons who tend to be better (or worse) than others at identifying some emotions, tend to be better (or worse) than others at identifying all emotions (at least all four of the emotions tested in the current study). Although, to this author's knowledge, the presence of a decoder by emotion interaction has not been tested for in the same context as it was here, these results lend some credence to parallel lines of emotion research that suggest some people are better at identifying emotions in general than are others (e.g., the emotional intelligence literature. See, Mayer, Salovey, & Caruso, 2008; Salovey & Mayer, 1989).

As discussed in the results section above, however, the non-significant decoder by emotion interaction effect should be interpreted with some degree of caution. The effect size for this non-significant test indicated that approximately a third of the variance could be accounted for by the interaction, indicating that the interaction might at least offer important information in how decoders identify emotions. To become more certain whether an interaction effect exists (or does not exist), the same analyses should be replicated with a larger sample. It should also be noted, however, that the statistic for effect size reported in the current study (partial η^2) was, in this case, a more liberal statistic than regular η^2, which suggested that only 16% of the variance was accounted for by the interaction. Therefore, it might actually be the case that the effect size, and not the *F* ratio, is the artifact of sample size.

It could be argued that generalizability of the decoder main effect is also in question due to such a small sample. Although the sample size is admittedly small in the current study, and might be of concern in the interaction effect, on the surface, generalizability does not appear to be a problem for the decoder main effect. Results from this study reflect results from comparable studies with much larger decoder sample sizes in the levels of identification accuracy for emotions in general and in the variance of individual decoders' identification accuracy (e.g., Juslin & Laukka, 2003; Scherer, 2003; Spackman, Brown, & Otto, 2009).

In addition to differences in decoder's identification accuracy levels, results also indicate that some emotions are more accurately identified than are others. This is not surprising. Differences across emotions in identification accuracy have been shown repeatedly in the literature over the years (see Scherer, 2003; Pittam & Scherer, 1993 for reviews). In general, however, most studies show high accuracy rates in identifying anger and sadness, moderate rates in identifying fear, and low rates in identifying happiness (see Scherer, 2003). In the present study, decoders were significantly more accurate in identifying anger and happiness than they were at identifying fear and sadness.

It should be noted that these differences might be due to the length of the recorded script. This issue will be discussed in more detail below but it should be noted that, in an effort to increase mundane realism, the script used in the current study was longer than scripts used in most other similar studies (see, Spackman, Brown, & Otto, 2009). It might be the case that it is more difficult to maintain compelling expressions of acted emotions over longer periods for some emotions (e.g., sadness) than it is for others (e.g., anger). If this were the case, it would make sense that identification accuracy rates in the current study might differ somewhat from those in other studies. Therefore, replicating this study with varied script lengths would be an important future step in this line of research.

In sum, three conclusions can be made from analyses of decoder accuracy. First, some decoders are more accurate at identifying emotion than are others. Second, some emotions are more accurately identified by decoders than are others. Third, decoders who are generally better (or worse) than others at identifying emotions tend to be better (or worse) than others at identifying all emotions.

The Nomothetic Nature of Emotion Decoding

Based on the expectation that there would exist a significant decoder by emotion interaction for scores of accuracy, the current study was interested in how such an interaction might be mediated by the four acoustic domains of frequency (F0), rate, intensity, and spectral characteristics. Although no evidence for an interaction effect was found, other important and interesting conclusions could be drawn from the results.

Because of recent evidence that individual encoders tend to express emotion vocally in varied ways, that is, idiographically, hypothesis one in the current study held that decoders would likewise focus on different acoustic properties when identifying emotions. Results did not provide evidence in support of this hypothesis. Instead, it appears that decoders exhibit remarkable consistency in their use of acoustic properties to identify emotion.

The level of this consistency is demonstrated in Figure 10, where the distinct U-shaped pattern of identified emotions plotted in the acoustic domain space is remarkably similar across all decoders. Specifically, results from a two-way ANOVA in hypothesis one indicate that, of the four acoustic property domains, wave frequency (F0) is the least utilized by decoders in identifying emotions. However, it is still important in differentiating between emotions. As displayed both graphically and statistically in Figures 5 though 9, each of the four domains is important in identifying at least one emotion. Specifically, decoders identified anger by primarily focusing on spectral characteristics, fear by primarily focusing on frequency (F0), happiness by primarily focusing on rate, and sadness by focusing on both intensity and rate. The consistency across persons displayed in Figure 10 is rare in these types of graphs and indicates a very high degree of similarity in the way decoders use acoustic properties to identify emotions. In other words, although there is evidence to believe that emotions are encoded idiographically, it appears as though they are decoded nomothetically, or similarly by everyone.

Caveats and Directions for Future Research

This statement should be made with some degree of caution, however. First, although recent research makes a compelling argument for the idiographicity of emotion encoding patterns, the majority of research in this area argues the opposite. Therefore, more research should be conducted to help make more certain that emotions are indeed encoded idiographically. Second, this is the first study to deeply investigate the acoustic nature of decoded emotions so results, compelling as they are, should be interpreted cautiously because not all potential confounds were accounted for. Below is a brief discussion of a few of these potential issues.

Script Length. As mentioned above, the script used in the encoding process of the current study was longer than scripts used in many other studies (see Spackman, Brown, & Otto, 2009). It is possible that encoders had a difficult time maintaining compelling emotion expressions through the entire duration of the relatively lengthy script. If this were indeed the case, it could be argued that only a portion of the total encoded script was actually representative of the intended emotion. This could at least partially explain why Spackman, Brown and Otto (2009) found idiographicity in encoded emotion expressions whereas most previous studies argue for nomotheticity in encoded emotion (see, e.g., Scherer, 2003).

More importantly, the script length could also explain the seemingly paradoxical relationship between the idiographic nature of encoded emotions and the nomothetic nature of decoded emotions. It is possible that decoders identified the expressed emotions primarily based upon cues from only snippets of the entire recording. For example, in one recording, a female gave a very brief giggle between two sentences of the script. At this moment, it became very clear she was expressing happiness. Had it not been for this giggle, it likely would have been more difficult for decoders to accurately identify the emotion in that recording.

Therefore, future research should evaluate the effect of script length in a number of ways. First, replicating the current study with varied script lengths could help to determine whether shorter scripts produce more nomothetic emotion expressions. Second, to better understand whether all or only a portion of the expression is used by decoders to identify an encoded emotion, it could be useful to create a task in which decoders identify portions of a recording that they find most useful to their decisions. This type of task is frequently used in both political science and advertising research, where participants are frequently requested to continuously rate the relative strength of some given arguments they are presented.

Familiarity. Another way to explain the seemingly paradoxical relationship between encoded and decoded emotions is to suggest there exists a general emotion decoding schema used by decoders to pick out only those elements of an emotion expression that are unique to a particular emotion. Such a schema would be adaptive in a sense because it would allow persons to quickly and accurately identify emotion expressions even though there potentially exist many ways to express emotion. However, it also makes sense to argue that as a person becomes more familiar with another person's particular emotion expression idiosyncrasies, the perceiver might stray from his or her general schema and utilize more idiographic methods appropriate to the emoter's particular way of expressing emotion. In other words, the degree to which a decoder is familiar with an encoder's particular method of emotion expression might affect the idiographicity or nomotheticity of decoding patterns. This is an issue not addressed by the current research and warrants future investigation.

Universality. As has been stated, the current research makes a strong case for the nomothetic nature of vocalized emotion decoding. At a deeper level, results from the research also make a strong case for the presence of distinct acoustic mediators for particular emotions. In other words, the current research begins to make clear which acoustic properties decoders find most important in differentiating between emotions.

One question not addressed in the current study, however, is the degree to which these profiles are universal across cultures. There exists a growing amount of evidence in the literature that emotion expressions can be fairly accurately identified cross-culturally (Juslin & Laukka, 2003). It would be interesting to learn whether persons from different cultures are not only similar in their levels of accurately identifying emotions, but also in the types of acoustic properties they employ to make such identifications. To make a definitive statement that vocalized emotions are, in fact, decoded nomothetically, they should be identified similarly by persons from differing backgrounds and cultures.

Similarly, because this study utilized a small number of decoders and only four encoded emotions, future research should test the degree to which these results are generalizable across other samples within the same culture as well as across other emotions.

Conclusion

In conclusion, this dissertation makes several significant contributions to vocal emotion expression research. It is among the first studies to deeply investigate how decoders use acoustic properties to identify vocalized emotions. Specifically, it makes a strong case that decoders employ a very consistent and precise method for identifying specific emotions.

However, these findings have even broader applicability. A better understanding of what people focus upon when identifying vocalized emotions could be of benefit to a number of industries. Perhaps most importantly, it has applicability to researchers and therapists in the areas of interpersonal relationships and intimacy. If, for example, one spouse is continuously taking offense to statements made by the other spouse that are not meant to be offensive, the other spouse could be trained to employ a tone of voice that utilizes fewer of the acoustic properties used to identify anger.

Of course, this is only one example of the applicability of the current research. Findings from the current study could also be employed in the development of compelling artificial emotion expressions, by drama coaches, or in forensic areas. The applied uses are, in fact, quite broad and the opportunity for expanded research in the area is great.

References

Aristotle & Hobbes, T. (1890). *Aristotle's Treatise on Rhetoric*. T. A. Buckley, Trans. London: Bell.

Arnold, M. B. (1960). *Emotion and personality*. New York: Columbia University Press.

Averill, J. (1980). A Constructivist view of emotion. In R. Plutchik & H. Kellerman (Eds.), *Emotion: Theory, research and experience, Vol. 1* (pp. 305-339). New York: Academic Press.

Bachorowski, J. A. (1999). Vocal expression and perception of emotion. *Current Directions in Psychological Science, 8*(2), 53-57.

Bachorowski, J. A., & Owren. M. J. (1995). Vocal expression of emotion: Acoustic properties of speech are associated with emotional intensity and context. *Psychological Science*, *6*(4), 219-224.

Bachorowski, J. A., & Owren, M. J. (2003). Sounds of emotion: Production and perception of affect-related vocal acoustics. *Annals of the New York Academy of Sciences, 1000*, 244-265. P. Ekman, J. J. Campos, R. J. Davidson, & F. B. M. de Waal (Eds)., Emotions inside out: 130 Years after Darwin's The Expression of the Emotions in Man and Animals.

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

Beal, D., Gillis, J., & Stewart, T. (1978). The lens model: Computational procedures and applications. *Perceptual and Motor Skills, 46*, 3-28.

Bernieri, F. J., Gillis, J. S., Davis, J. M., & Grahe, J. E. (1996). Dyad rapport and the accuracy of its judgment across situations: A lens model analysis. *Journal of Personality and Social Psychology, 71*(1). 110-129.

Boersma, P., & Weenink, D. (2005). Praat (Version 4.4.07). Amsterdam, Netherlands: University of Amsterdam

Brunswik, E. (1952). *The conceptual framework of psychology*. Chicago: Chicago University Press.

Brunswik, E. (1956). *Perception and the representative design of experiments*. Berkeley, CA: University of California Press.

Cicero, M. T. (2001). *De Oratore*. Oxford University Press. New York.

Cooksey, R. W. (1996). *Judgment analysis: Theory, methods, and applications.* San Diego, CA: Academic Press.

Cornelius, R. R. (1996). *The science of emotion. Research and tradition in the psychology of emotion.* Upper Saddle River, NJ: Prentice Hall.

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

de Gelder, B., Teunisse, J. P., & Benson, P. J. (1997). Categorical perception of facial expressions and their internal structure. *Cognition & Emotion, 11,* 1–23.

de Gelder, B., & Vroomen, J. (1996). Categorical perception of emotional speech. *Journal of the Acoustic Society of America*, *100*, 2818.

Eibl-Eibesfeldt, I. (1989). *Human ethology.* New York: Aldine.

Ekman, P. (1972). Universal and cultural differences in facial expressions of emotions. In J. K. Cole (Ed.), *Nebraska symposium on motivation*, 1971 (pp. 207-283).

Ekman, P. (1992). An argument for basic emotions. *Cognition & Emotion*, *6*, 169-200.

Ekman, P. (1994). Strong evidence for universals in facial expressions: A reply to Russell's mistaken critique. *Psychological Bulletin, 115*(2), 268-287.

Ekman, P. (1999) Basic emotions. In T. Dalgleish and T. Power (Eds.) *The handbook of* *cognition and emotion*. Pp. 45-60. New York.: John Wiley & Sons.

Ekman, P. & Davidson, R. J. (1994). *The nature of emotion: Fundamental questions.* New York: Oxford University Press.

Ekman, P. & Friesen, W. V. (1978). *Facial action coding system*. Palo Alto, CA: Consulting Psychologists Press.

Ekman, P., Friesen, W. V., & Ellsworth, P. (1982). What emotion categories or dimensions can observers judge from facial behavior? In P. Ekman (Ed.), *Emotion in the human face* (pp. 39-55). New York: Cambridge University Press.

Ekman, P., Friesen, W. V., O'Sullivan, M., Chan, A., Diacoynni-Tarlatzis, I., Heider, K., Krause, R., LeCompte, W. A., Pitcairn, T., Ricci-Bitti, P. E., Scherer, K. R., Tomita, M., & Tzavaras, A. (1987). Universals and cultural differences in the judgments of facial expressions of emotion. *Journal of Personality and Social Psychology, 53*(4), 712-717.

Ekman, P., Sorenson, E. R., & Friesen, W. V. (1969). Pan-cultural elements in the facial displays of emotion. *Science*, *164*, 86-88.

Elfenbein, H., Marsh, A., & Ambady, N. (2002). Emotional intelligence and the recognition of emotion from facial expressions. In L. Feldman-Marrett & P. Salovey (Eds.), *The wisdom in feeling* (pp. 37-59). New York: Guilford Press.

Etcoff, N. L., & Magee, J. J. (1992). Categorical perception of facial expressions. *Cognition, 44*, 227–240.

Flom, R. & Bahrick, L. E. (2007). The development of infant discrimination of affect in multimodal and unimodal stimulation: The role of intersensory redundancy. *Developmental Psychology, 43*(1), 238-252.

Frijda, N. H. (1986). *The Emotions.* New York: Cambridge University Press.

Frijda, N. H. (1993). Moods, emotion episodes, and emotions. In M. Lewis & J. M. Haviland (Eds.) *Handbook of Emotions* (pp. 381-404). New York: Guilford Press.

Gray, J. A. (1985). A whole and its parts: Behaviour, the brain, cognition and emotion. *Bulletin of the British Psychological Society, 38*. 99-112.

Hohmann, G. W. (1966). Some effects of spinal cord lesions on experienced emotional feelings. *Psychophysiology, 3*, 143-156.

Howell, D. C. (2002). *Statistical methods for psychology,5th Ed.* Belmont, CA: Thompson Wadsworth.

Izard, C. (1971). *The face of emotion*. Appleton-Century-Crofts, New York.

Izard, C. (1977). *Human emotions*. New York: Plenum Press.

Izard, C. (1980). Cross-cultural perspectives on emotion and emotion communication. In H. C. Triandis & W. Lonner (Eds.), *Handbook of cross-cultural psychology* (pp. 185-221). Boston: Allyn & Bacon.

James, W. (1884). What is an emotion? *Mind, 9*, 188-205.

Juslin, P. N. (1997). Perceived emotion expression in synthesized performances of a short melody: Capturing the listener's judgment policy. *Musicae Scientiae, 1*, 225-256.

Juslin, P. N. (2000). Cue utilization in communication of emotion in music performance: Relating performance to perception. *Journal of Experimental Psychology: Human Perception and Performance, 26,* 1797-1813.

Juslin, P. N. & Madison, G. (1999). The role of timing patterns in recognition of emotional expression from musical performance. *Music Perception, 17*, 197-221.

Juslin, P., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin, 129*, 770-814.

Lange, C. G., & James, W. (1922). *The emotions*. Baltimore: Williams and Wilkins.

Lazarus, R. S. (1991). *Emotion and adaptation*. New York: Oxford University Press.

LeDoux, J. E. (1986). The neurobiology of emotion. In J. E. LeDoux & W. Hirst (Eds.), *Mind and brain: Dialogues in cognitive neuroscience* (pp. 207-232). Cambridge: Cambridge University Press.

LeDoux, J. E. (1987). Emotion. In F. Plum & V. B. Mountcastle (Eds.), *Handbook of physiology. The nervous system: Vol. 5. Higher function* (pp. 419-459). Washington, DC: American Physiological Society.

Lieberman, P., & Michaels, S. B. (1962). Some aspects of fundamental frequency and envelope amplitude as related to the emotional content of speech. *Journal of the Acoustical Society of America, 34*, 922-927.

Mandler, G. (1975). *Mind and emotion.* New York: Wiley.

Mayer, J. D., Salovey, P. & Caruso, D. R. (2008). Emotional Intelligence: New ability or eclectic traits, *American Psychologist, 63*(6), 503-517.

McDougall, W. (1926). An *introduction to social psychology*. Boston: Luce.

Mowrer, O. H. (1960). *Learning theory and behavior*. New York: Wiley.

Murray, I., & Arnott, J. (1993). Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion. *Journal of the Acoustic Society of America, 93*, 1097-1108.

Oatley, K. (2004). *Emotions: A brief history*. Malden, MA: Blackwell.

Oatley, K., & Johnson-Laird, P. N. (1987). Towards a cognitive theory of emotions. *Cognition & Emotion, 1*, 29-50.

Ortony, A. , Clore, G. L., & Collins, A. (1988). *The Cognitive Nature of Emotions*. New York, Cambridge University Press.

Ortony, A., & Turner, T. J. (1990). What's basic about basic emotions? *Psychological Review*, 97, 315-331.

Panksepp, J. (1982). Toward a general psychobiological theory of emotions. *The Behavioral and Brain Sciences, 5*, 407-467.

Papez, J. W. (1937). The brain considered as an organ: neural systems and central levels of organization. *American Journal of Psychology, 49*, 217-232.

Pittam, J., & Scherer, K. (1993). Vocal expression and communication of emotion. In M. Lewis & J. M. Haviland (Eds.), *Handbook of emotions*. New York: Guilford Press.

Plutchik, R. (1980). A general psychoevolutionary theory of emotion. In R. Plutchik & H. Kellerman (Eds.), Emotion: *Theory, research, and experience: Vol. 1. Theories of emotion* (pp. 3-33). New York: Academic.

Rencher, A. C., & Scott, D. T. (1990). Assessing the contribution of individual variables following rejection of a multivariate hypothesis. *Communication in Statistics: Simulation and Computation*, 19(2), 535-553.

Roseman, I. J. (1984). Cognitive determinants of emotion. In P. Shaver (Ed.), *Review of personality and social psychology:* Vol. 5. *Emotions, relationships and health* (pp. 11-36). Beverly Hills, CA: Sage.

Russell, J. A. (1994). Is there universal recognition of emotion from facial expression? A review of cross-cultural studies. *Psychological Bulletin, 115*, 102-141.

Russell, J. A. (1995). Facial expressions of emotion: What lies beyond minimal universality? *Psychological Bulletin, 118*(3), 379-391.

Salovey, P., & Mayer, J. D. (1989). Emotional intelligence. *Imagination, Cognition, and*

Personality, 9(3), 185-211.

Schachter, S., & Singer, J. E. (1962). Cognitive, social, and physiological determinants of emotional state. *Psychological Review, 69*, 121-128.

Scherer, K. R. (1982). Emotion as a process: Function, origin, and regulation. *Social Science Information, 21*, 555-570.

Scherer, K. R. (1986). Vocal affect expression: a review and a model for future research. *Psychological Bulletin, 99*, 143-165.

Scherer, K. R. (1989). Vocal measurement of emotion. In R. Plutchik, & H. Kellerman (Eds.) *The measurement of emotions.* (pp. 233-259) San Diego, CA: Academic Press.

Scherer, K. R. (1993). Neuroscience projections to current debates in emotion psychology. *Cognition & Emotion, 7*, 1-41.

Scherer, K. R. (2000). Psychological models of emotion. In: J. Borod (Ed.), *The Neuropsychology of Emotion*. Oxford University Press, Oxford/New York, 137-162

Scherer, K. R. (2003). Vocal communication of emotions: A review of research paradigms. *Speech Communication, 40*, 227-256.

Scherer, K., Banse, R., Wallbott, H., & Goldbeck, T. (1991). Vocal cues in emotion encoding and decoding. *Motivation and Emotion, 15*, 123-148.

Scherer, K., Schorr, A., & Johnstone, T. (Eds.) (2001). *Appraisal processes in emotion: Theory, methods, research.* Oxford University Press, New York, Oxford.

Scherer, K. R. & Oshinsky, J. S. (1977). Cue utilization in emotion attribution from auditory stimuli. *Motivation and Emotion, 15*, 123-148.

Solomon, R. C. (1993). *The passions: Emotions and the meaning of life.* Indianapolis, IN: Hackett.

Spackman, M., Brown, B., & Otto, S. (2009). Do emotions have distinct vocal profiles? A study of idiographic patterns of expression. *Cognition & Emotion, 23(8)*, 1565-1588.

Stewart, T. (2004, November 18-19). *Notes on a form of the lens model equation for logistic regression analysis*. Paper presented at the Brunswik Society Meeting, Minneapolis, MN.

Swets J. A., Tanner W. P., & Birdsall T. G. (1961). Decision Processes in Perception. *Psychological Review, 68(5)*, 301-340.

Tomkins, S. S., (1962). *Affect, imagery, consciousness. The positive affects vol.1*. Springer, New York.

Tomkins, S. S. (1984). Affect theory. In K. R. Scherer & P. Ekman (Eds.), *Approaches to emotion* (pp. 163-195). Hillsdale, NJ: Erlbaum.

Van Bezooijan, R. (1984). *The characteristics and recognizability of vocal expressions of emotions*. Dordrecht, Netherlands: Foris.

Wallbott, H., & Scherer, K. (1986). Cues and channels in emotion recognition. *Journal of Personality and Social Psychology, 51*, 690-699.

Watson, J. B. (1930). *Behaviorism*. Chicago: University of Chicago Press.

Weiner, B., & Graham, S. (1984). An attributional approach to emotional development. In C. E. Izard, J. Kagan, & R. B. Zajonc (Eds.), *Emotions, cognition, and behavior* (pp. 167-191). New York: Cambridge University Press.

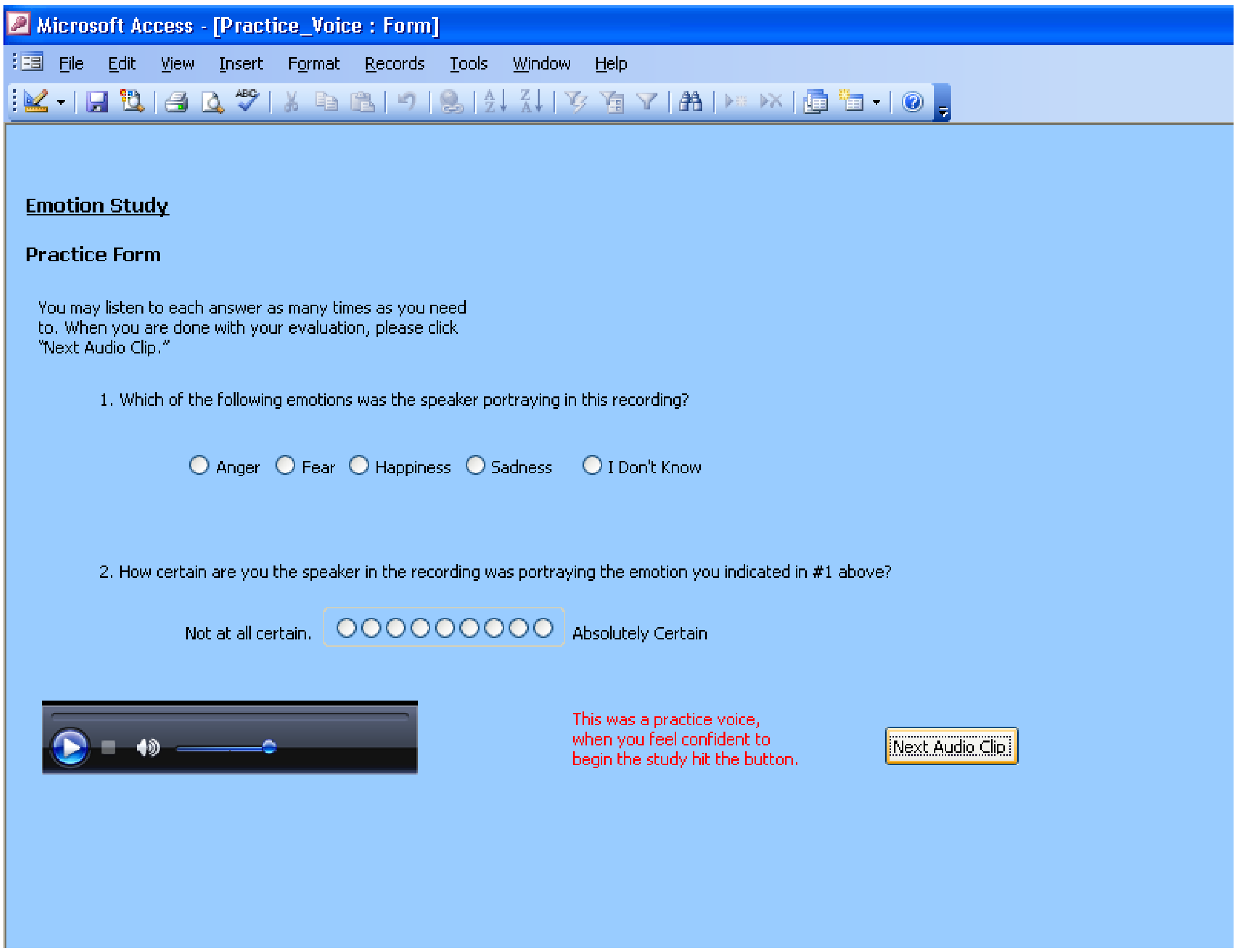
Wickens, T. D. (2001). *Elementary Signal Detection Theory*. Oxford/New York: Oxford University Press.

Wundt, W. (1874/1905). *Grundzüge der physiologischen Psychologie, [Fundamentals of physiological psychology, orig. pub 1874] fifth ed.* Engelmann, Leipzig.

Appendices

Appendix A

Screen shot of Access response application



Appendix B

Identification ratios calculated for the ANOVA in hypothesis one

Individual

			Identification Ratios			
Decoder	**Gender**	**Emotion**	**F0**	**Intensity**	**Rate**	**Spectral**
1	Male	Anger	.079	.185	.099	.300
1	Male	Fear	.058	.039	.207	.117
1	Male	Happy	.037	.200	.395	.009
1	Male	Sad	.100	.258	.231	.131
2	Male	Anger	.090	.170	.107	.367
2	Male	Fear	.241	.058	.090	.176
2	Male	Happy	.049	.126	.255	.062
2	Male	Sad	.128	.274	.292	.127
3	Male	Anger	.093	.174	.106	.241
3	Male	Fear	.104	.068	.090	.205
3	Male	Happy	.019	.121	.287	.030
3	Male	Sad	.014	.246	.311	.140
4	Male	Anger	.093	.134	.065	.305
4	Male	Fear	.144	.095	.088	.202
4	Male	Happy	.025	.165	.379	.037
4	Male	Sad	.114	.416	.486	.188
5	Female	Anger	.074	.113	.091	.185
5	Female	Fear	.190	.051	.088	.196
5	Female	Happy	.037	.118	.237	.056
5	Female	Sad	.068	.291	.345	.159
6	Female	Anger	.050	.103	.075	.282
6	Female	Fear	.134	.077	.172	.166
6	Female	Happy	.027	.134	.285	.011
6	Female	Sad	.051	.211	.274	.141
7	Female	Anger	.073	.127	.074	.286
7	Female	Fear	.145	.085	.165	.185
7	Female	Happy	.048	.146	.283	.017
7	Female	Sad	.081	.281	.280	.124
8	Female	Anger	.149	.165	.035	.522
8	Female	Fear	.137	.027	.196	.235
8	Female	Happy	.045	.122	.246	.038
8	Female	Sad	.169	.434	.461	.173

Group

			Identification Ratios			
Decoder	**Gender**	**Emotion**	**F0**	**Intensity**	**Rate**	**Spectral**
All	All	Anger	.085	.144	.085	.293
All	All	Fear	.146	.061	.138	.181
All	All	Happy	.036	.140	.291	.033
All	All	Sad	.084	.288	.322	.145

Appendix C

ANOVA summary table for two-way ANOVA on identification ratios in hypothesis one

Dependent Variable: Identification Ratio

Source	Type III SS	df	MS	F	*p*	Partial η^2
Corrected Model	.279(a)	7	0.04	3.97	.001	.170
Intercept	1.393	1	1.39	138.69	.000	.505
IndGrp	0.000	1	0.00	0.02	.882	.000
Acoustic	0.109	3	0.04	3.61	.015	.074
IndGrp * Acoustic	0.000	3	0.00	0.00	1.00	.000
Error	1.366	136	0.01			
Total	5.242	144				
Corrected Total	1.645	143				

Appendix D

d' Statistics submitted to the ANOVA in hypothesis two

			d' Statistics			
Decoder	**Gender**	**B/W**	**Anger**	**Fear**	**Happy**	**Sad**
1	Male	best	1.889	1.702	1.487	2.220
1	Male	worst	3.779	1.206	1.729	1.691
2	Male	best	2.023	1.942	2.209	1.872
2	Male	worst	2.220	1.855	2.619	1.577
3	Male	best	2.684	2.037	2.355	1.487
3	Male	worst	2.684	1.415	2.526	1.642
4	Male	best	1.691	1.215	2.144	1.577
4	Male	worst	2.050	.860	2.023	1.308
5	Female	best	3.068	1.574	4.773	1.629
5	Female	worst	2.406	1.064	2.533	1.562
6	Female	best	1.889	1.574	1.729	2.209
6	Female	worst	2.526	1.702	3.068	1.702
7	Female	best	1.702	1.377	1.456	1.131
7	Female	worst	2.711	1.853	3.266	1.642
8	Female	best	1.534	.897	1.747	.662
8	Female	worst	1.215	1.415	2.057	1.150

Appendix E

Results from Tukey's HSD post hoc analysis on *d'* statistics.

(I) Decoders 1 through 8	(J) Decoders 1 through 8	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
1	2	-.07675	.271849	1.00	-.95735	.80385
	3	-.14087	.271849	.999	-1.02147	.73972
	4	.35438	.271849	.891	-.52622	1.23497
	5	-.36325	.271849	.878	-1.24385	.51735
	6	-.08700	.271849	1.00	-.96760	.79360
	7	.07062	.271849	1.00	-.80997	.95122
	8	.62825	.271849	.319	-.25235	1.50885
2	1	.07675	.271849	1.00	-.80385	.95735
	3	-.06412	.271849	1.00	-.94472	.81647
	4	.43113	.271849	.755	-.44947	1.31172
	5	-.28650	.271849	.962	-1.16710	.59410
	6	-.01025	.271849	1.00	-.89085	.87035
	7	.14737	.271849	.999	-.73322	1.02797
	8	.70500	.271849	.195	-.17560	1.58560
3	1	.14087	.271849	.999	-.73972	1.02147
	2	.06412	.271849	1.00	-.81647	.94472
	4	.49525	.271849	.611	-.38535	1.37585
	5	-.22238	.271849	.991	-1.10297	.65822
	6	.05387	.271849	1.00	-.82672	.93447
	7	.21150	.271849	.993	-.66910	1.09210
	8	.76913	.271849	.123	-.11147	1.64972
4	1	-.35438	.271849	.891	-1.23497	.52622
	2	-.43113	.271849	.755	-1.31172	.44947
	3	-.49525	.271849	.611	-1.37585	.38535
	5	-.71762	.271849	.179	-1.59822	.16297
	6	-.44138	.271849	.733	-1.32197	.43922
	7	-.28375	.271849	.964	-1.16435	.59685
	8	.27388	.271849	.970	-.60672	1.15447
5	1	.36325	.271849	.878	-.51735	1.24385
	2	.28650	.271849	.962	-.59410	1.16710
	3	.22238	.271849	.991	-.65822	1.10297
	4	.71762	.271849	.179	-.16297	1.59822
	6	.27625	.271849	.968	-.60435	1.15685
	7	.43387	.271849	.749	-.44672	1.31447
	8	**.99150***	**.271849**	**.019**	**.11090**	**1.87210**
6	1	.08700	.271849	1.00	-.79360	.96760
	2	.01025	.271849	1.00	-.87035	.89085
	3	-.05387	.271849	1.00	-.93447	.82672
	4	.44138	.271849	.733	-.43922	1.32197
	5	-.27625	.271849	.968	-1.15685	.60435
	7	.15762	.271849	.999	-.72297	1.03822
	8	.71525	.271849	.182	-.16535	1.59585
7	1	-.07062	.271849	1.00	-.95122	.80997
	2	-.14737	.271849	.999	-1.02797	.73322
	3	-.21150	.271849	.993	-1.09210	.66910
	4	.28375	.271849	.964	-.59685	1.16435
	5	-.43387	.271849	.749	-1.31447	.44672
	6	-.15762	.271849	.999	-1.03822	.72297
	8	.55763	.271849	.466	-.32297	1.43822
8	1	-.62825	.271849	.319	-1.50885	.25235
	2	-.70500	.271849	.195	-1.58560	.17560
	3	-.76913	.271849	.123	-1.64972	.11147
	4	-.27388	.271849	.970	-1.15447	.60672
	5	**-.99150***	**.271849**	**.019**	**-1.87210**	**-.11090**
	6	-.71525	.271849	.182	-1.59585	.16535
	7	-.55763	.271849	.466	-1.43822	.32297

Appendix F

Identification ratios entered into the MANOVA model for hypothesis three

			Identification Ratios							
			Best				Worst			
Decoder	**Gender**	**Emotion**	**F0**	**Intensity**	**Rate**	**Spectral**	**F0**	**Intensity**	**Rate**	**Spectral**
1	M	Anger	.149	.270	.155	.366	.018	.120	.282	.190
1	M	Fear	.028	.004	.059	.185	.110	.107	.490	.042
1	M	Happy	.000	.146	.394	.000	.164	.267	.513	.075
1	M	Sad	.116	.128	.198	.036	.088	.445	.382	.302
2	M	Anger	.183	.190	.173	.425	.041	.180	.164	.328
2	M	Fear	.182	.023	.000	.190	.300	.123	.329	.131
2	M	Happy	.043	.067	.228	.077	.144	.204	.481	.098
2	M	Sad	.110	.180	.240	.027	.132	.414	.405	.268
3	M	Anger	.124	.188	.152	.260	.068	.196	.127	.240
3	M	Fear	.062	.012	.002	.151	.176	.177	.324	.304
3	M	Happy	.001	.089	.209	.029	.092	.198	.480	.116
3	M	Sad	.015	.191	.279	.128	.012	.329	.430	.160
4	M	Anger	.134	.215	.112	.412	.058	.115	.216	.226
4	M	Fear	.000	.014	.069	.172	.563	.250	.308	.278
4	M	Happy	.000	.091	.263	.025	.106	.271	.598	.098
4	M	Sad	.108	.257	.331	.115	.117	.634	.709	.306
5	F	Anger	.123	.140	.108	.196	.043	.124	.210	.164
5	F	Fear	.197	.000	.005	.084	.210	.217	.438	.273
5	F	Happy	.008	.071	.162	.022	.127	.194	.417	.110
5	F	Sad	.122	.303	.253	.113	.025	.306	.546	.165
6	F	Anger	.086	.182	.091	.378	.019	.100	.272	.222
6	F	Fear	.000	.025	.045	.189	.291	.164	.500	.148
6	F	Happy	.000	.094	.297	.000	.116	.180	.390	.089
6	F	Sad	.081	.152	.260	.118	.036	.303	.405	.179
7	F	Anger	.156	.220	.126	.452	.031	.096	.231	.175
7	F	Fear	.079	.000	.071	.174	.214	.157	.385	.218
7	F	Happy	.026	.111	.328	.004	.112	.197	.448	.078
7	F	Sad	.142	.299	.228	.124	.049	.282	.442	.118
8	F	Anger	.231	.235	.173	.611	.075	.125	.136	.389
8	F	Fear	.170	.000	.063	.326	.119	.094	.349	.159
8	F	Happy	.016	.080	.211	.000	.161	.196	.431	.137
8	F	Sad	.289	.418	.337	.040	.092	.451	.475	.253

Note: Identification ratios with a value of .000 were initially very close to zero, but slightly negative. This mathematical impossibility was an artifact of Stewart's (2004) correction of Cooksey's (1996) correction to the Brunswikian lens model. Values were rounded to zero to enable calculation of the MANOVA model.

Appendix G

Summary tables for univariate analyses of identification ratios on the four acoustic domains and of two measures of accuracy

Source	df	Type III SS	MS	F	p
F0					
gender	1	0.000	0.000	0	0.996
decoder(gender)	6	0.050	0.008	0.73	0.629
emotion	3	0.087	0.029	2.56	0.072
gender*emotion	3	0.002	0.001	0.07	0.974
emotion*decoder(gender)	18	0.066	0.004	0.33	0.993
Intensity					
gender	1	0.005	0.005	0.46	0.504
decoder(gender)	6	0.028	0.005	0.41	0.864
emotion	3	0.463	0.154	**13.47**	**<.001**
gender*emotion	3	0.002	0.001	0.06	0.982
emotion*decoder(gender)	18	0.080	0.004	0.39	0.981
Rate					
gender	1	0.001	0.001	0.04	0.847
decoder(gender)	6	0.037	0.006	0.21	0.972
emotion	3	0.506	0.169	**5.65**	**0.003**
gender*emotion	3	0.018	0.006	0.2	0.894
emotion*decoder(gender)	18	0.093	0.005	0.17	1
Spectral Characteristics					
gender	1	0.000	0.000	0	0.949
decoder(gender)	6	0.057	0.010	0.97	0.462
emotion	3	0.534	0.178	**18.09**	**<.001**
gender*emotion	3	0.006	0.002	0.2	0.898
emotion*decoder(gender)	18	0.101	0.006	0.57	0.895
Phi Coefficient					
gender	1	0.001	0.002	0.17	0.681
decoder(gender)	6	0.327	0.055	**5.81**	**<.001**
emotion	3	0.442	0.147	**15.66**	**<.001**
gender*emotion	3	0.038	0.013	1.34	0.279
emotion*decoder(gender)	18	0.197	0.011	1.17	0.342
d prime					
gender	1	0.013	0.013	0.04	0.838
decoder(gender)	6	0.363	0.894	**3.02**	**0.019**
emotion	3	0.943	0.314	**11.21**	**<.001**
gender*emotion	3	0.227	0.409	1.38	0.266
emotion*decoder(gender)	18	0.697	0.261	0.88	0.601