The Effect of Pitch on the Creation of Emotional Meaning in Music and Language

Aimee Siebert

Bethel College

In partial fulfillment of PSY 482: Psychology Seminar and COA 430: Communication Arts Seminar

Dwight Krehbiel, Paul Lewis, John McCabe-Juhnke & M.E. Yeager

Abstract

Music and language are two human media known to communicate emotion. Burgeoning research comparing music and non-verbal language has identified acoustic characteristics, like pitch, that both media share. This study seeks to determine whether pitch functions similarly in music and language to communicate emotion. Participants listen to four actors' readings of the same Shakespearean monologue and to eight other sound files: a derived prosody file and a transcribed music file for each of the four monologues, for a total of 12 sound files. This produces four sets of three sound files that preserve the pitch movements of the actor's voice in three types of sound, yielding stimuli that can be directly compared for pitch's effect on a listener's perception of emotion in different communication media. Emotion is measured in two response forms: participants' subjective ratings and physiological recordings. Results show that participants' ratings of activation and efficiency of emotional communication are preserved across the three communication media, suggesting that pitch differences from the four actors' readings influence these ratings for music and language. Other findings indicate that speech stimuli generate the strongest emotional ratings of the three media types. Results for activation also corroborate past literature which shows women have stronger responses to emotional communication than men. Discussion covers how the importance of activation in this study may be due to the focus on the emotion of anger in the stimuli to which participants listened.

The Effect of Pitch on the Creation of Emotional Meaning in Music and Language

Effective communication comes from more than just words. Most anyone can defend this claim with anecdotal experience in which, for example, the response "I'm fine" means dramatically different things depending on the tone of voice the speaker uses. The same words convey different meanings depending on where, how and with whom they are spoken. More support for extra-verbal communication structure is the human experience of music. Oliver Sacks in says that "we humans are a musical species no less than a linguistic one" (p. 1). Most people experience profound emotional connection to some kind of music: a kind of connection that aches, raises the hair across the arms and neck, and rejuvenates when little else could. Music also seems to mean something. It is a brand of communication. These phenomena suggest that human communication, though increased in precision of meaning and valence by words, produces some level of meaning though nonverbal cues and channels. Understanding the mechanism of creating meaning nonverbally is potentially useful to manifold areas of music and communication, including, but not limited to, public speaking, music composition, music and language education, neurological research, music and speech/language therapy and artificial intelligence. Beyond complex research and practical applications, this question also engages us at a natural level; as users of these two communication types and as creatures sensitive to emotion, how can we help but be interested in the ways we create meaning via these media?

Though extrinsic elements of communication like speaker-audience relationship and location are undoubtedly influential on meaning, the thrust of this research is aimed at meaning created by more intrinsic, nonverbal characteristics of two human media of communication: speech and music. Specifically, literature was examined concerning the effect of pitch on emotional meaning derived from nonverbal speech and music. Relevant literature to this study can be broken into several subcategories and sub-subcategories to articulate the depth and breadth of scholarship that supports the comparability of pitch's contributions to emotion content in speech and music. They include the following:

- Communication Frameworks
- Affect, Emotion, & Valence Studies
    - Prosody & Emotion
    - Music & Emotion
- Music/Language Comparison Literature (which can be examined specifically with)
    - Neurological/Physiological Arguments
    - Interrelated Effects of Music and Language

Literature detailing methodologies for exploring music/language interactions and influential for the procedure of this research are described in the Methods section. Reviewing this literature also informs the relevance of the original experiment conducted on pitch's contribution to emotional meaning and response in speech and music.

Literature Review

Communication Frameworks

In their text *Pragmatics of Human Communication,* Watzlawick, Beavin & Jackson (1967) outlined their interactional perspective of communication, in which all messages contain two dimensions: content and relational. Roughly, the content-based dimension answers the question "what?" and includes the actual words and exact phrasing of a message, characteristics of the communication that Watzlawick, Beavin & Jackson also refer to as digital communication. Their relational-based dimension of communication answers "how?" in terms of how the exact message is conveyed and ought to be understood, and roughly corresponds with Watzlawick, Beavin & Jackson's idea of analog communication, which is any nonverbal form. This is the dimension of communication with which music and language comparison studies are most concerned. Watzlawick, Beavin, & Jackson's second axiom of the interactional perspective describes this relational dimension of communication as "metacommunication" because it communicates about communication. In addition to external characteristics of communication mentioned in the introduction, emotional expression falls into the category of relational communication. It informs the communicators about the other communicators involved in the message, whether a message should be taken seriously or in jest, and cues appropriate responses depending on that information.

Emotional communication also contributes to how genuine a message is perceived to be and consequently, how invested listeners become in a message. Petty & Cacioppo's (1986) work with the Elaboration Likelihood Model for persuasion models how investment in a message will affect communication. According to this model, "Under conditions of high elaborative likelihood, attitudes are most affected by argument quality. Under conditions of low elaborative likelihood, attitudes are most affected by peripheral cues" (Petty & Cacioppo, 135). In other words, when a communicator is highly invested in a topic, she is more likely to attend to evidence for multi-sided arguments, but in low levels of investment, peripheral cues like how pleasant a speaker is, or how many point she makes will be important to persuading an audience. There are many other peripheral cues from which the peripheral route of the Elaboration Likelihood Model could benefit, but if all variations in a set of communication stimuli other than the voice of the speaker were eliminated, the emotional expression of each speaker would be the independent variable that affects audience responses.

Wolfe & Powell (2006) assert that gender also contributes to how individuals understand emotional communication. When examining expressions of dissatisfaction among mixed-gender student work groups, Wolfe & Powell disproved the stereotype that women complain more than men, but indicated rather that the genders complain for different reasons. Women are more likely to be making an indirect request for action by complaining, whereas men express dissatisfaction to excuse behavior or make themselves seem superior. But for both genders, emotional communication adds another layer to the meaning of what is exactly being said.

Emotion, Valence and Affect Studies

Emotion—the experience of it and the effective communication of it—is central to human experience and successful social encounters. Those who struggle with emotional expression or understanding also struggle to fit into society, often to a pathological degree, as in the cases of some types of schizophrenia, autism and other mental disorders which are characterized by flat affect. Emotions, like motives, serve an activating and directing role for behavior. Emotions are evolutionarily-maintained heuristics that help us decide what to do in response to external stimuli as much, if not more, than logic does (Nolen-Hoeksema, Fredrickson, Loftus & Wagenaar, 2009).

Classical emotion models like those of James (1890/1950), Schachter & Singer (1962) Lazarus (1991), and Rosenberg (1998) all describe emotion not as a static state, but as a process with components. For Lazarus and Rosenberg, the person's relationship with his or her environment moderates his or her cognitive appraisal of a certain event, including whether or not it was personally relevant. Based on this cognitive appraisal, the person would have the subjective experience of a particular emotion, thought-action tendencies related to the emotion, and internal bodily changes associated with the emotion. These internal experiential and physiological responses to a particular emotion would lead to more visible behavioral responses to emotion. Mauss & Robinson (2009) and Barrett (2006) elaborated on the relationships of these three components of emotion in their own models. Mauss & Robinson argue that there can be no gold standard for measuring emotional response, and measures accessing the three components: 1. subjective experience 2. physiological change, and 3. behavior are equally relevant and do not seem to be interchangeable.

Schachter & Singer (1962) developed their model of emotion in which the presence of a stimulus creates general physical arousal, of which the person must form a cognitive appraisal in order to reach a subjective experience of a particular emotion. This contrasts with the James-Lange theory (James 1890/1950) in which the stimulus causes a physiological arousal pattern specific to a particular emotion, and that arousal pattern alone is enough to cause the subjective experience of an emotion. Schachter and Singer's theory may coincide better with experience because it allows for the "misattribution of arousal" where someone mistakes arousal caused by an innocuous source (i.e. adrenaline rush standing on a high bridge) as an emotion (falling in love with the person next to you) (Dutton & Aron, 1974).

While not all arousal is a sign of emotion, most emotion does cause some level of arousal. The stronger the emotional arousal, the stronger the physiological responses: for instance, the sympathetic nervous system in response to highly arousing stimuli causes increases in blood pressure, heart rate, perspiration, and respiration rate. Blood is also diverted from the internal organs to the brain and skeletal muscles in preparation for action. Research has shown that some individuals are more sensitive to these physiological changes than others are, or, in other words, have heightened interoceptive sensitivity. These arousal-focused individuals emphasize feelings of arousal more in their emotion reports over time than non-arousal focused individuals (Barrett, Bliss-Moreau, Quigley & Aronson, 2004).

According to Barrett (2006) and her meta-analysis of emotion literature, arousal is one of two major dimensions that make up affective experience. The other is valence or how pleasant an emotion is. Valence, according to Barrett derives from the process of valuation, where something is judged as helpful or harmful. Based on this meta-analysis, Barrett formed an affective circumplex with arousal and pleasantness as the two axes. Just as people differ in the extent to which they are arousal-focused, they differ in valence focus too (Barrett, 2004; Barrett, 2005).

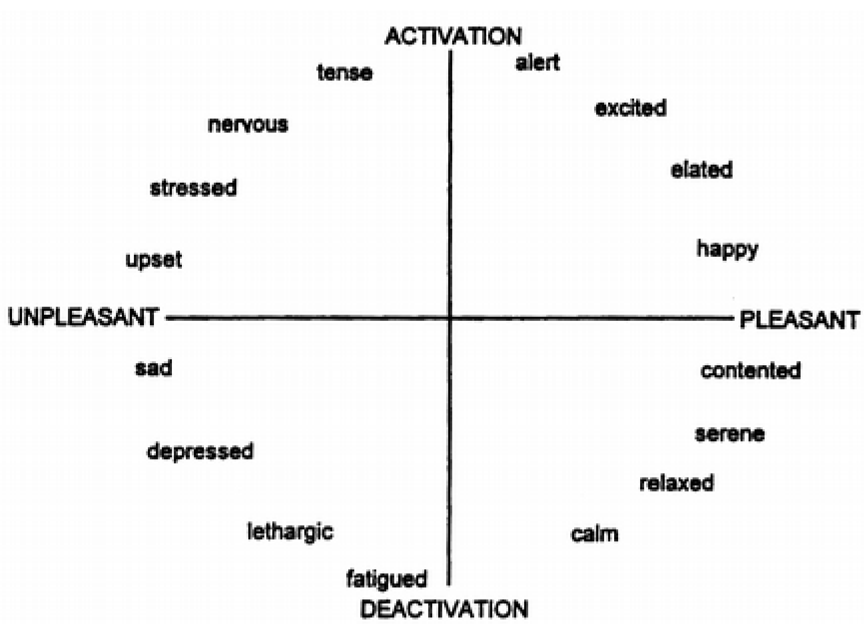

Figure 1 – Barrett's arousal and valence circumplex (http://psycnet.apa.org/journals/psp/81/4/images/psp_81_4_684_fig1a.gif)

Positive emotions—those on the right side of Barrett's circumplex—have shown innumerable beneficial effects. Fredrickson (2000, 2002) developed the broaden-and-build theory, which argues that positive emotions cause the way people think and act to broaden, which in turn would build lasting personal resources that the person might not otherwise have encountered. Consequently, they are more complex, resilient people. Negative emotions, however, are also highly adaptive in threatening situations, in which their narrowing and focusing effect allows people to zero in on threats and deal with them decisively.

Gender plays a stable role in an individual's degree of emotional awareness. Barrett, Lane, Sechrest, and Schwartz (1999) showed that women consistently score higher on an emotional awareness performance test and display more complexity and differentiation in their articulation of emotional experiences than men. These robust findings remain even when controlled for age, scholastic performance, socioeconomic status, culture and verbal intelligence. Unfortunately, this high degree of emotional awareness might be corrupted in the stereotype that women are the more emotional sex. What Barrett and Bliss-Moreau (2009) found, however, is that this judgment is based more on explanations of behavior than on behavior itself. In their experiment, participants, even when given situational information, more frequently judged female targets depicting emotions as "emotional" whereas men would be judged as "having a bad day" (Barrett & Bliss-Moreau, 649).

Recent emotion research has also studied the relationship of affect and cognition. According to Duncan and Barrett (2007), the distinction between the two mind constructs does not hold up in neural mechanisms. Affect has direct, simultaneous effect on sensory processing, which signals what visual sensations stand for in the present and how to act on them in the future (Duncan & Barrett, 2007; Barrett & Bar, 2009). It also appears affect is needed for normal conscious experience, language fluency and memory (Duncan & Barrett, 2007).

Prosody and emotion. Links between the function of prosody—the rhythm, stress and intonation of speech—and emotion expression/perception, first recognized a long time ago, are becoming more and more apparent in current literature (Herman, 2006; Patterson & Johnsrude, 2008; Pittam & Scherer; 1993; Fortenbaugh, 1986). Both Pittam & Scherer (1993) and

Fortenbaugh (1986) allude to Greek thinkers who believed prosody affected the expression of emotion, both real and faked, and exhibited social influence on interpersonal interactions. Aristotle, in his discussion of delivery, said "voice is an important medium for conveying character," "a speaker's delivery helps make discourse not only clear and enjoyable, but also persuasive" and discussed how variation in voice helped to distinguish one speech act from another (Fortenbaugh, 1986, pp. 244, 246). Charles Darwin found that voice carries affective signals (Pittam & Scherer, 1993). Ann K. Wennerstrom (2001), according to David Herman (2006), identified several affectively-related functions of prosody, including a grouping function of lexical and syntactical elements, which cues turn-taking in interpersonal conversation, similar to Aristotle's evaluation. She also noted prosody's function in indicating contrasting relationships, and expression of emotion. Patterson & Johnsrude (2008) experimentally demonstrated that prosody could convey non-linguistic information on size, sex, background, social status & the emotional status of the speaker. Mulac & Giles's (1996) found that how old you sound best predicts negative psychological judgments. It seems that we, as a society, like the sound of young, lively voices better than older ones.

This interrelatedness of prosody and emotion should be expected considering the physiological effects of affective arousal on speech-production organs (Oudeyer, 2003; Scherer, 1986; Steeneken & Hansen, 1999; Pittam & Scherer, 1993). Steeneken & Hansen (1999) studied military personnel under situations of stress and found respiratory changes and increased muscle tension in the vocal cords, which changed the quality of speech, particularly in terms of pitch, intensity, duration, and the spectral envelope. In addition to respiration and muscle tension, changes have also been detected in a speech phonation and articulation due to characteristic physiological responses of different emotion states (Pittam & Scherer, 1993; Scherer, 1986).

Oudeyer (2003) utilized these predictable effects of certain emotions' physiological states on speech, especially on pitch, timing and voice quality, to develop algorithms that allow robots to express emotions. Oudeyer found that these algorithms produced robotic emotions that humans can identify with similar accuracy to emotion expression by humans, which hovers around 66% across cultures and emotions (Scherer, Banse, & Wallbott, 2001; Pittam & Scherer, 1993). Greater error in emotion identification seems to occur when compared emotions have similar valence or arousal levels (Mullenix et al., 2002; Oudeyer, 2003; Pittam & Scherer, 1993), which suggests that Barrett's findings about people's sensitivity to arousal and valence are indeed emotionally relevant.

Markel, Bein & Phillis (1973) also contributed to this body of research on predictable physiological effects on speech for particular emotions with their finding of normative relationships between content and tone-of-voice for given emotions. When people talk about an affectively charged subject, certain voice qualities are expected to coincide depending on the emotion being expressed. Scherer, Ladd & Silverman (1984) determined that there were particular intonational variables which contributed to affect only in interaction with grammatical features of message content, whereas others, like voice quality and the fundamental frequency of a person's voice can convey affective information independently of verbal content. Mino (1996) confirmed these findings in a practical setting, where in a simulated employment interview, content and vocal cues provided different information that informed different responses. In Mino's study, vocal delivery is associated with assertiveness, enthusiasm, emotional stability, sincerity and outgoingness, characteristics that are not unrelated to Addington (1971) and Black's (1971) measures of speaker's competence, trustworthiness and dynamism. Mino also found that the combination of good content and good delivery were found in the employer's number one candidates. Interestingly, poor content and good delivery applicants were rated second most interesting candidates, but that combination was also correlated with the least sincere scores. This shows the partial independence of voice and content variables, and also the social preference for dynamic voices, even at the expense of sincerity. Petty & Cacioppo’s (1986) Elaboration Likelihood Model predictions for low levels of investment might be relevant to this last finding, since employers were not realistically choosing employees, and therefore wouldn’t be highly invested in content-based information. Lab settings might be particularly disposed to low levels of investment for participants.

Burgoon, Blair & Strom (2008) showed the importance of verbal and nonverbal interaction in their study too. Their participants were given access to verbal transcripts, verbal transcripts with voiced recordings or verbal transcripts with audio/visual recordings of a truthful or deceptive subject. Vocal cues in the second two conditions increased participants' ratings of the subject's completeness, honesty, clarity, relevance, dominance and credibility. The best discrimination and detection of deception also took place when vocal cues were available.

Prosody research has also shown a gender interaction with prosodic perception; women are generally found to be more sensitive to prosodic cues than men (Besson, Magne & Schon, 2002; Scherer, Banse & Wallbott, 2001). It is important to note that women are generally more sensitive to emotion expression, so conceptualizing prosody as a form of emotion expression corresponds to these separate findings.

Other emotional prosody research has zeroed in on specific acoustic variables that correlate with certain emotions. Addington (1971) and Pearce (1971) showed the effect of vocal delivery, particularly the patterns of pitch in vocal delivery, on listener's judgment of the speaker's competence, trustworthiness and dynamism. In both studies, higher and more variable pitch was associated with dynamism, while lower pitch, pitch & rate agreement, reduced inflection range, less volume, and articulation were associated with feelings of trustworthiness and competence. Black (1942) had also correlated certain prosodic variables with preference for a speaker's voice, including greater total and functional pitch range, greater number of upward inflections and greater extent of downward inflections. Combining Addington, Pearce and Black's work, it would seem that we prefer more dynamic voices.

Oudeyer's (2003) review of computer-based techniques of sound manipulation indicated that the pitch (F0) contour, intensity contour, and timing of utterances in speech are the most salient aspects of speech that reflect emotion. Dellaert, Polzin & Waibel (1996), in their study of four basic emotions (happiness, sadness, anger and fear), identified seven global statistics of pitch signal relevant to emotion perception: 1. mean pitch, 2. standard deviation, 3. minimum, 4. maximum, 5. range, 6. slope and 7. speaking rate.

Of those four basic emotions, much acoustic research has been done on anger specifically, and pitch has been found to play a large role in its communication. Mullennix et al. (2002) also investigated the effects of angry emotional tone, and though their content was only a word long, they showed that the fundamental frequency (F0) contour (a common measure of pitch) appears to remain steady or fall slightly and the mean duration is shorter for an 'angry' word, which corroborates other research they had consulted. Oudeyer's computer manipulation of emotion indicated that anger is correlated with high mean pitch and pitch variance, little variation in phoneme durations, fast rhythm, unaccented final syllables, and falling pitch contours for all syllables. Pittam & Scherer (1993) corroborate Oudeyer's and Mullennix's findings, also finding anger associated with high mean pitch and F0 variability, high articulation rate, and increased numbers of downward directed F0 contours.

Scherer's (1986) investigation of vocal affect expression performed an extensive meta-analysis on existing research and discovered several relationships between acoustic variables and anger, though he became concerned that different kinds of anger were being studied, e.g. cold and hot anger. He made sure to differentiate between these types in his own research, and his rage/hot anger is what is most relevant to this study. J. Darby found that anger exhibits a high level, a wide range, and a large variability in pitch, as well as loud volume and fast tempo (as cited in Scherer, 1986). These variables are more relevant to arousal than to valence, and Scherer's finding was that anger's degree of pleasantness is very open to individual experiences. Scherer studied the same global statistics as Dellaert, Polzin and Waibel, but supplemented them with other acoustic variables like F0 perturbation, F1 mean, Formant bandwidth/precision, intensity mean/range/variability, frequency range, high frequency energy and spectral noise. For hot anger, Scherer found that it exhibited narrow hedonic valence, very tense activation, and extremely full power. What these characteristics translated to in terms of acoustic variables was much greater F0 range, F0 variability, mean intensity, and high-frequency energy; decreased F0 shift regularity; greater F1 mean, intensity range, intensity variability, and frequency range; much smaller F1 bandwidth, lower F2 mean, and increased formant precision (p. 158). What's more, Scherer found that the main effects of these variables had a conspicuous lack of interactions, indicating that they are all relevant to affective expression.

Finally, there is research to suggest that these variations in vocal affect expression are hard-wired in the brain. Frick (1985) found that emotion is encoded and decoded with a high degree of agreement across cultures. This would make sense if all humans shared a brain structure that mediated the use of these emotional vocal expressions, which is what Frick found in the anterior cingulated cortex, a brain structure that's activated when these vocal expressions are used at will to communicate.

Music and emotion. Perhaps even more familiar than the effects of prosody on emotion, are the tangible effects of music on emotion. Patrick N. Juslin and John A. Sloboda (2001) assert that given the strength of music's relationship to emotion "emotional aspects of music should thus be at the very heart of musical science" (p. 4). Juslin and Sloboda identify several opposites through which the relationship between emotion and music can be understood. Are emotional responses to music a product of biology or of culture? Do we perceive the emotion of the other person or have emotion induced within ourselves? Is emotion private experience or public expression, and is emotion separate from a musical experience or does it rather "create" musical experience? Does music has intrinsic properties that "induce" or "force" emotion in the listener, or does the listener "[use] the music as a resource in a more active process of emotional construction" (p. 453). It is clear from Juslin and Sloboda's anthology that both sides of these pairs of opposites contribute to the emotional effects of music. Of these theoretical dichotomies used to approach the subject of emotion in music, the last one is most impactful for this current research. Its debate, intrinsic vs. extrinsic sources of emotional responses to music, is not unlike the theories of communication that range from simple theories with the sender of a unidirectional clear message through a channel to a receiver, to complex models where meaning is created by both communicators through continuous feedback from each other and their context. There is truth and effectiveness in both kinds of communication theories as well.

The literature on which this project focuses uses the theory of intrinsic properties in the music which induce emotion in the listener. Findings demonstrating music's ability to elicit deep and significant emotion are robust. Sloboda & Juslin have found behavioral, physiological and experiential components of emotion elicited by music in experiments that involve self-reports, behavioral measures like decision time, distance approximation, and writing speed, as well as physiological reactions (Juslin & Sloboda, 2001, p. 84). Juslin (1997b and 1997a, 2000) as cited in Juslin & Sloboda (2001), showed that these varied reactions were not necessarily incidental because his studies demonstrated that listeners could accurately decode emotional meanings 75 percent of the time in a forced-choice format (four times higher than chance) and that professional music performers can communicate emotions accurately to listeners.

Sloboda (1992) theorized music's emotive qualities offer access to and intensification/release of existing emotions, as well as an alternative perspective on emotion. His research identified structural features of music that elicited physiological responses like crying/a lump in the throat, spine shivers/goosebumps, and racing heart/pit of the stomach sensations, which are indicative of emotional experience.

Correlations between structures of music and emotional responses suggest that people have expectations for certain musical events in a piece, and temporal presentation (on time, early or late) affects a listener's emotional responses. Lerdahl and Jackendoff (1983) thought these expectations formed a musical grammar that we all develop. In their text "A Generative Theory of Tonal Music" they explain how musical grammar, which includes pitch-related aspects like "being in a key" creates meaning in real time, including moments of indeterminacy when expectations are delayed or not met. Musical affect, according to Lerdahl and Jackendoff, is wrapped up in this musical expectancy and remains unchanged in spite of familiarity because the musical grammar does not change. Palmer (1992) conceptualized this musical grammar as a culture's shared mental representations for musical knowledge which are the means by which we communicate musical ideas and emotions, perform music, perceive it and comprehend it.

Shaffer (1992) saw it as a play of tension and relaxation over different musical forms, and Steinbeis & Koelsch (2008) showed that violations of harmonic tension resolution patterns produced two event-related potentials: N400 and ERAN, that are traditionally related to violations of semantic meaning in language. Music's ability to produce these same event-related potentials seems to indicate that tension and relaxation of musical expectancies also have semantic values that inform music's emotional meaning to listeners.

As in prosody, gender influences emotion perception and expression in music. O'Neill (1997) found that girls have higher positive attitudes toward music at all ages and they give more favorable ratings while listening to music. Crozier (1997) noticed the effect of gender identity in his study of conformity concerning musical tastes. For Crozier, gender forms one of many possible social communities which endorse certain preferences for music, and musical perception is related to those social identities. Collectively, this research might suggest a society's development of musical expectancies for internal features of music that also produce affective responses.

Much music and emotion literature overlaps with prosody by focusing on features of music that have analogs in language. Different researchers all or some of these dimensions and call them different things, but overall, there appear to be three major dimensions of music: pitch, rhythm, and timbre, that influence emotion perception and expression in music. Juslin & Sloboda (2001) call these properties of music like metre, rhythm, tonality, etc. "representational" because they "are central to the recognition, identification, and performance of music" (p. 4) and their book *Music and Emotion* focuses on how these representational processes are related to affective processes. Kellaris & Kent (1993) called their three main factors tonality, tempo and texture, and they measured the effect of orthogonal changes to these factors on participant's reports of emotional dimensions like pleasantness, arousal and surprise. They found that tonality change affects pleasantness and surprise, tempo affects arousal and pleasure, and texture moderates the effects of tonality and tempo. Alpert & Alpert (as cited in Kellaris & Kent, 1993) seemed to be manipulating these relationships between tempo, tonality, and pleasantness to induce happy and sad moods by fast, major music, and slow, minor music respectively. Bruner (1990) also found that excitement is associated with major modalities in music, fast and medium range pitch, syncopated rhythm, dissonant harmony, and loud volume. He also found that there seems to be a moderate level of arousal (or excitement) that people prefer to feel, and they select music accordingly. When participants in Bruner's experiment were angered by the experimenter before listening to the music, they subsequently selected and preferred music of less complexity and tempo, which are variables of arousal in music. Bruner also found that moderate complexity correlated with higher liking of ads and probability of purchase. Bruner thought that in his experiment, music was acting as a moderator or amplifier of aroused emotion. Like Frederickson, he also noted that using music to induce negative moods prompted individuals to use deliberate analytical processing of a situation, while positive moods led to the use of heuristics.

Juslin compiled emotional data from music that he organized into a circumplex on valence and arousal axes like Barrett’s (2004). He identified the properties of music associated with five emotions: tenderness (positive valence, low activity), happiness (high valence, high activity), sadness (low activity, negative valence), and anger and fear, both of which are associated with high activity and negative valence. Anger, which is of interest to the present study, is correlated with musical qualities like high sound level, sharp timbre, spectral noise, fast mean tempo, small tempo variability, staccato articulation, abrupt tone attacks, sharp duration contrasts, accents on unstable notes, large vibrato extent, and no ritardando (Juslin & Sloboda, 2001, p. 315). Another metaanalysis of properties of musical structure was compiled by Alf Gabrielsson and Erik Lindstrom (as shown in Juslin & Sloboda, 2001, p. 235-239). They identify similar properties with anger, including a sharp amplitude envelope, staccato articulation, complex/dissonant harmony, loudness, upward pitch contour, minor mode, high pitch level, small pitch variation, complex rhythm, fast tempo, many harmonics in timbre, sharp timbre, and atonality.

These metaanalyses suggests that people have musical expectations for particular emotions. Kellaris & Kent found consumption-related results in which congruity between the mood of the music and a product in an advertisement produced more positive purchase intent. This means sad music would (and did) encourage consumers to purchase "Missing you" cards better than happy music. This may be a musical expression of the normative relationships as Markel, Bein & Phillis (1973) found in tone-of-voice and emotion well as demonstration of behavioral effects of music-elicited emotion. Kellaris & Kent recommended that another step in this research would be to "manipulate tonality and hold speed constant to avoid confounding pleasant feelings with arousal" (1993, p. 396).

Pitch, tempo, and timbre elements in music also interact with emotion and verbal language much the same way as prosody. Like emotional responses to language and other factors, emotional responses to music have three levels: autonomic, denotative and interpretive (Wieczorkowska et al., 2005) which correspond roughly to the physiological, behavioral and experiential levels found in other studies of emotion. Allan (2006) found that pop music, presented with original lyrics, altered lyrics or only instrumentals, caused different advertising effects. The music presented with lyrics compared to without, produced stronger attention and memory effects. The strength of Allan's findings was moderated by the personal significance of the music to the listener, which may be linked to Zhu & Meyers-Levy's (2005) finding that different demands on processing resources affected the kinds of meaning to which music listeners were attentive. According to them, music contains both referential meaning and embodied meaning. Referential meaning is context-dependent meaning associated with external world concepts, whereas embodied meaning is "purely hedonic, context-independent, and based on the degree of stimulation the musical sound affords" (2005, p. 333). Zhu & Meyers-Levy found that non-intensive processing engages neither of these meanings, while demands on few processing resources cause listeners to be sensitive to referential meaning. Embodied meaning is only salient when listeners are devoting large amounts of processing resources to attending to the music. These findings may be a music-specific expression of Petty & Cacioppo's Elaboration Likelihood Model.

Like in prosody, Lee, Skoe, Kraus & Ashley (2009) found that individuals who have been musically trained develop greater sensitivity to certain affective elements of music. In their study, musicians had heightened subcortical brain responses to particular harmonics and to some complex combinatorial sounds. It seems that the mechanism underlying perception of musical harmony is also more precise in musicians and correlated to their years of musical training.

Music and Language Comparison Literature

Even across the separate treatments of language and music, common relationships to emotion for the two domains are clear, but the act of deliberately comparing responses to music and with those to language within the same study is a flourishing enterprise. Juslin describes the rising functionalist perspective of music which holds that “music performers are able to communicate emotions to listeners by using the same acoustic code as is used in vocal expression of emotion" (Juslin & Sloboda, 2001, p. 321). More and more researchers are applying empirical methodologies and analyses to music and language events to better understand the evident overlaps between the two communication media. Findings have yielded neurological/physiological correlates, interactive therapeutic effects and other shared characteristics relevant to the range of meanings produced by music and language. A small percentage of those findings are clarified below.

Neurological/physiological arguments. Auditory features are among the first variables we receive in communication and are subsequently processed by the brain, and much research indicates that it is this encoding level is shared in speech and music. Above and beyond the effects musical training has on musical sensitivity, Kraus, Skoe, Parbery-Clark & Ashley (2009) and Strait, Kraus, Skoe & Ashley (2009) were able to show that musical experience enhances perception of emotion in all sound at the subcortical level seen by Lee et al. (2009) in purely musical studies. Strait sees the potential in musical training for "boosting deficient (neurological) mechanisms" which would "strengthen bonds between people and systems within individual brains" (Ferdinand, 2009, p. 2).

Research from the same lab as Strait was able to show that length of musical training also produces more efficient and enhanced brainstem responses to the most complex parts of sound, which are the parts of sound that patients with language disorders struggle with (Wong, Skoe, Russo, Dees & Kraus, 2007). These strengthened effects were found even when the individuals were not paying attention to the sound (i.e. when they were given a different task to focus on) and were related to the ability to phase-lock to stimulus periodicity, an ability which requires perception of pitch. In other words, participants perceived and encoded pitch at brainstem levels even when their attention was not focused on the sound. Subcortical encoding and processing of frequency and temporal features of sound were also enhanced by audiovisual presentations for musically-trained participants (Musacchia, Sams, Skoe & Kraus, 2007). These subcortical responses might be the mechanism for enhanced detection of deception when acoustic cues are available, as seen in Burgoon, Blair & Strom (2008) research. Musacchia, Strait & Kraus (2008) furthered this line of research by showing that early brainstem responses were subsequently related to early cortical response timing peaks further along in brain processing of sound. Musacchia, Strait & Kraus predicted that this early timing and neural representations of pitch, timing and timbre are shaped in a coordinated manner for both language and music. Koelsch et al. (as cited in Patel, 2008) also measured event-related potentials shared between music and speech and showed that they did not differ in the time course, strength or neural generators of N400, a semantically related peak. These studies suggest that emotion is encoded faster for individuals with musical training and that this encoding is pertinent to both speech and music messages, perhaps explain musicians' higher language-learning abilities.

Zatorre & Gandour (2008) found hemispheric specializations for aspects of sound that nonetheless spanned language, music and other auditory domains. It seems the right hemisphere is involved in pitch processing irrespective of domain. This does not negate the well-supported finding that speech is better processed by the left hemisphere, but Zatorre & Gandour's finding was that this left hemispheric processing was connected to intelligibility and therefore to phonetic and semantic patterns from memory. This further supports the idea that some meaning encoding happens at a lower level than verbal meaning, and it is at this level that music and language may share acoustic features and neurological resources.

Interactive effects of music and language. Some of the emotional effects resulting from music have been hypothesized to be due to the resemblance of musical features to prosodic features relevant to the same emotion (Juslin & Sloboda, 2001; Shaffer, 1992). Curtis & Bharucha (2009) have conclusively shown that the same minor third interval that expresses sadness in music, communicates the same emotion in speech. "These findings support the theory that human vocal expressions and music share an acoustic code for communicating sadness" (p. 1) and perhaps other emotions. On a more interactive level, Alter & Knosche (2003) found that people break speech and song into auditory phrases through the same markers: boundary tones, prefinal lengthening and pause insertion. Stegemoller et al. (2008) studied the greater energy at frequency ratios associated with the 12-tone music scale, and found that greater musical experience caused the individual's voice to utilize less energy at frequency ratios not associate with those 12 tones, which may indicate an ability of musicians to better align their speaking and singing voices. Ross et al (2007) predicted that all humans would have a sense of tonality that would develop preferences for those specific tonal intervals.

Speech/Language Therapy and Music Therapy are used to treat a range of disorders and deficits. The literature defends the positive effects of these therapies in a wide range of measures, from well-being to emotion identification/understanding, to increased participation in social settings like the classroom, for individuals with a wide range of deficits or disorders (Geist et al., 2008; Spackman et al., 2005; Magee et al., 2006). Where the literature becomes particularly compelling for this study is the instances where individuals with language deficits show marked benefits from music therapy above and beyond the benefits they experienced from speech/language therapy. Geist et al. (2008) performed the case study of a four-year old with global development delay who showed increased engagement in the classroom due to the use of music therapy in addition to the prescribed speech/language therapy. Spackman et al. (2005) performed an emotion study with facial expressions and musical expression of emotion, which indicated that the ability to identify even nonverbal expressions of emotion (like the music and facial expressions) is closely entwined with language development and impairment. It is worth pondering whether the ability to name an emotion affects one's experience of it. Magee et al. (2006) showed that music therapy improves linguistic prosody and phonation, a finding corroborated by dozens of recent studies which show that musical training/experience improves not only sensitivity to emotion in music but in language as well, likely by means of the neurological circuits described above (Strait et al., 2009; Thompson, Schellenberg & Husain, 2004; Stegemoller et al, 2008). Schon, Magne & Besson's work (2004) might have clarified the significant element of emotion perception in their findings that music training facilitates and enhances pitch contour processing in both music and language. Musicians are sensitive to weaker fundamental frequency variations and show shorter onset latency to brain potentials that are equally strong to clearer frequency variations.

Patel et al. (1998) investigated the shared effects of music and language from the other direction. They studied individuals with amusia, a neurological deficit in processing pitch and musical memory and recognition, and compared their prosodic and musical discrimination abilities to control participants. The processing deficits were shown to be variable by individual, but the level of performance for the amusia participants was statistically similar across the language and music domains, which further suggests shared neural resources for prosody and music. However in his work with individuals who had difficulties with both music and language syntax, he found that they did not struggle with perceiving pitch patterns or short-term memory for tones, indicating a separate acoustical path for these elements of music and language.

Having consulted the references addressed in the literature review, and planning the measurements outlined in the following Methodology section, it is clear that pitch elements appear in both music and language and influence the perception of emotion in each domain. Therefore, this study seeks to add to the available literature by holding other variables equal and answer whether pitch elements operate to the same degree or in the same fashion in both media. The following primary and secondary hypotheses have been formed.

Primary Hypothesis 1: Sound files (speech, prosody, music) derived from the same actor

Participants will respond to the set of three sound files (speech, prosody and music) derived from an actor with similar subjective ratings of emotion and preference, as well as with similar physiological responses.

Primary Hypothesis 2: Sound files (speech, prosody, music) derived from the same actor

The strength of ratings and physiological responses will be strongest in the speech condition, where emotional meanings are clarified by words.

Secondary Hypothesis 1: Effects of Musical Training and Gender

As found in past studies, women and more musically trained individuals will be more sensitive to and exhibit stronger responses to emotion in all three types of sound files, in all three types of emotional measures.

Secondary Hypothesis 2: Responses to Particular Acoustic Variables

Pitch variability (range) and average pitch will be most closely correlated with participant's preferences, due to their importance perception of anger and dynamism in past studies (Scherer, 1986; Addington, 1971; Pearce, 1971). More specifically, the closer the actor's voice matches the cluster of pitch variables identified for hot anger by Scherer (1986), the more preferred that interpretation will be, particularly for speech, following what the normative relationship Markel et al. (1973) found between voice and content depending on the emotion being expressed. People expect a certain 'tone-of-voice' for a particular emotion.

Methods

Participants

35 participants (15 males, 20 females; age = 18-23 yrs., mean = 19.74 yrs.) from the Bethel student body were solicited from psychology and philosophy classes and received extra credit for participating.

Acoustic Stimuli and Design

The goal in stimuli selection and creation was to eliminate all variation in content, necessitating the use of the same monologue for the base of all the sound files, so as to isolate pitch features as independent variables influencing emotional dependent variables. The monologue selected was Shylock's "I am a Jew" speech from Shakespeare's *The Merchant of Venice* (see Appendix 1). This seminar used recordings from the professional performances of Shylock by Al Pacino (*The Merchant of Venice*, 2004, Spice Factory) and Orson Welles (*The Merchant of Venice*, 1969) as well as the competitive amateur performances by Adam Brown and Paul Olivier Bros at the English Speaking Union's 2007 and 2009 National Shakespeare Competitions, respectively, as the designated "speech" stimuli.

Each of these speech stimuli were filtered for lowpass at 250 Hz and 26 dB using Audacity (Mazzoni, 2010) to extend Pearce's (1971) methodology for eliminating intelligibility of speech and producing content-free "prosody" stimuli for their studies.

A recently developed open source package called Praat (Boersma, 2009) performs acoustic analysis and sound manipulation and has a program called Prosogram v2.4f (2009), which yields an adjusted readout of the pitch contour of a person's voice. These adjusted pitch contours allegedly account for the thresholds at which human perception notices a difference in pitch, which raw pitch contours neglect. These prosograms are read in semitone intervals. These semitone intervals were be transcribed into Finale composition software (2009, Make Music, Inc.,) and turned into pure tones of music for “music” stimuli.

This results in a 3 Media X 4 Performers set of 12 stimuli to which all participants were exposed, making this experiment a repeated-measures, within-subjects design.

Apparatus and Procedure

Concerning the collection of emotional data relevant to psychological (self-report) and physiological responses of emotion, this seminar modeled past Bethel psychology of music experiments and utilized the ActiveTwo Data Acquisition System (BioSemi, Amsterdam, Netherlands), powered by a DC battery pack via active Ag/AgCl electrodes (MettingVanRijn. Kuiper, Dankers & Grimbergen, 1996) to record peripheral physiological responses to stimuli. These physiological responses are related especially to the arousal dimension of emotion and include heart rate, galvanized skin response (GSR), temperature and facial muscle movements (EMG). The signals were saved with the use of LabVIEW-based ActiVIEW software (BioSemi, Amsterdam, Netherlands).

Participants’ experiential emotional responses of valence and arousal were recorded post-listening periods using the Self Assessment Manikin (SAM; Lang, Bradley, & Cuthbert, 1999). The SAM instrument has been shown to have strong reliability coefficients for valence and arousal (Cronbach’s alpha, range = 0.83 - 0.93; Jennings, McGinnis, Lovejoy & Stirling, 2000) and it has been effectively used in music research (Morris & Boone, 1998). Its application here to prosodic and speech-based stimuli should be appropriate if my hypothesis concerning the functional relationship between music and language is strong. Post-listening ratings of liking and efficiency (“how effective was the piece in conveying emotion”) were gathered by participants moving a slider along a simple 9-point Likert scale on the same LabVIEW VI to indicate their responses.

When participants arrived, they were seated in front of a computer on which they would make their experiential emotional ratings and hooked up to the ActiveTwo apparatus. Once hooked up, participants were given instructions about the experiment. They were told for each sound file they listened to, there would be a minute-long baseline, a listening episode in which they might listen to three different types of acoustical stimuli: prosody, music and speech, and then a rating period in which they would be asked to make several ratings about each piece. These questions of pleasantness and arousal asked not how the participants felt about listening to the stimuli, but asked them to describe the emotions they believed the creator of the sound was expressing. They were shown pictures of the SAM rating scale and how to use it. They were also asked how much they like the stimuli and how efficient the stimuli were at expressing the creator's emotion. It was explained that before, during and after the listening periods, physiological data would be recorded from the sensors I had attached to their body.

First, baseline measures of participants' mood the day of the session were solicited via the SAM rating scale before the listening session began. Then they listened to a practice piece and gave the ratings they would use during the listening sessions. At this point, a pause was taken for any questions the participants had, and then they proceeded with the 12 acoustic stimuli. The four prosody files were always presented first (randomized internally), then music, then speech, to separate the presentation of the speech and prosody pieces. It was the hope that this would further decrease intelligibility of the prosody pieces by presenting them first and at a distance from their particular speech pieces. This grouping of speech files also encouraged participants to compare within medium rather than across media, but the four files of each medium would be randomly presented to avoid order effects. After the completion of the experiment, each participant filled out a short debriefing sheet with demographics including gender and musical training, and any concerns or questions were addressed.

Data Analysis

Data for heart rate, EMG, GSR and temperature were divided into separate sound files and processed in a LabVIEW-VI to generate second-by-second averages. From those averages, an average for a five-second baseline period was taken. The first 96 seconds of data from each sound file were then processed as derivations from that baseline average. The average derivation across the sound file was used as the statistically-tested measures of heart-rate, EMG, GSR and temperature for each participant for each song. These physiological data were entered with the psychological, and demographic data in a consistent order in EXCEL.

At a later point, acoustical properties of the sound files derived from each of the four performers was added to the data for testing. The acoustical data investigated are duration (as a rough measure of tempo), minimum, maximum, range, mean and standard deviation of pitch, and mean absolute slope as given by Praat. These are basic measures that were relevant to several studies in the literature (Pearce, 1971; Black, 1942; Mullennix et al, 2002; Ververidis & Kotropoulos, 2006; Scherer; 1986, Oudeyer, 2003; Steeneken & Hansen, 1999; Scherer, Ladd & Silverman; 1984). Using these stimuli allow for statistical comparison of this study's acoustical stimuli to one another and to expected acoustic parameter patterns for the expression of anger (Oudeyer, 2003; Pittam & Scherer; 1993) as well as those patterns associated with ratings of speaker credibility and voice preference (Black, 1942; Pearce, 1971; Addington, 1971). It may be that some of these musical structure properties are associated with listener preference or ratings of efficiency.

Relationships amongst the types of emotional responses were tested using correlations and Hierarchical Linear Modeling.

To test the primary hypotheses, the five experiential measures of emotion: activation, pleasantness, efficiency and liking, as well as pertinent physiological averages, would each be subjected to a two-way, repeated-measures ANOVA. This test could answer whether the speech files had stronger emotional responses than the other types of sound files, or whether performer had a significant effect on any emotional ratings or measures.

Hierarchical Linear Modeling could test these same relationships while controlling for the two mediating factors of gender and musical training or other demographic/debriefing data, such as whether a participant's ability to understand the prosody files affected their emotional ratings and measures.

Results

Primary Hypothesis 1

The hypothesis that participants would respond to three types of sound derived from the same performer with similar emotional responses, both psychological and physiological, would indicate that emotional responses are influenced by pitch patterns in the performers' voices which stay constant across the three media. Results did not show this relationship universally, but for particular measures.

Measures of activation and efficiency differed significantly by performer. Numerical summaries for both variables are shown by performer in Tables 1 and 2. For Activation (Table 1), the values are inverted so that higher values are actually lower activation.

	mean	sd	0%	25%	50%	75%	100%	n
Al	2.771200	2.447857	0.00	0.79	2.43	4.066	9.66	105
Adam	2.440571	2.170317	0.00	0.83	2.07	3.430	9.49	105
Paul	4.607905	2.361031	0.12	3.12	3.94	6.000	10.00	105
Orson	6.266286	2.607073	0.73	4.16	7.04	8.380	10.00	105

Table 1. Activation

	mean	sd	0%	25%	50%	75%	100%	n
Al	6.546571	2.518774	0	5.00	6.59	8.61	10	105
Adam	6.655429	2.534690	0	5.49	7.02	8.63	10	105
Paul	5.198667	2.606787	0	3.91	5.30	7.10	10	105
Orson	5.178000	2.477552	0	3.33	5.62	6.89	10	105

Table 2. Efficiency

A two-way repeated-measures analysis of variance showed main effects of performer and medium on participants' ratings of activation, but no interaction effects, suggesting the effects are additive and independent of one another.

Error: Within

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Medium	2	134.13	67.07	14.280	1.027e-06 ***
Performer	3	287.35	95.78	20.394	2.616e-12 ***
Medium:Performer	6	44.16	7.36	1.567	0.1554
Residuals	396	1859.84	4.70		

Figure 2 shows these relationships in graphical form.

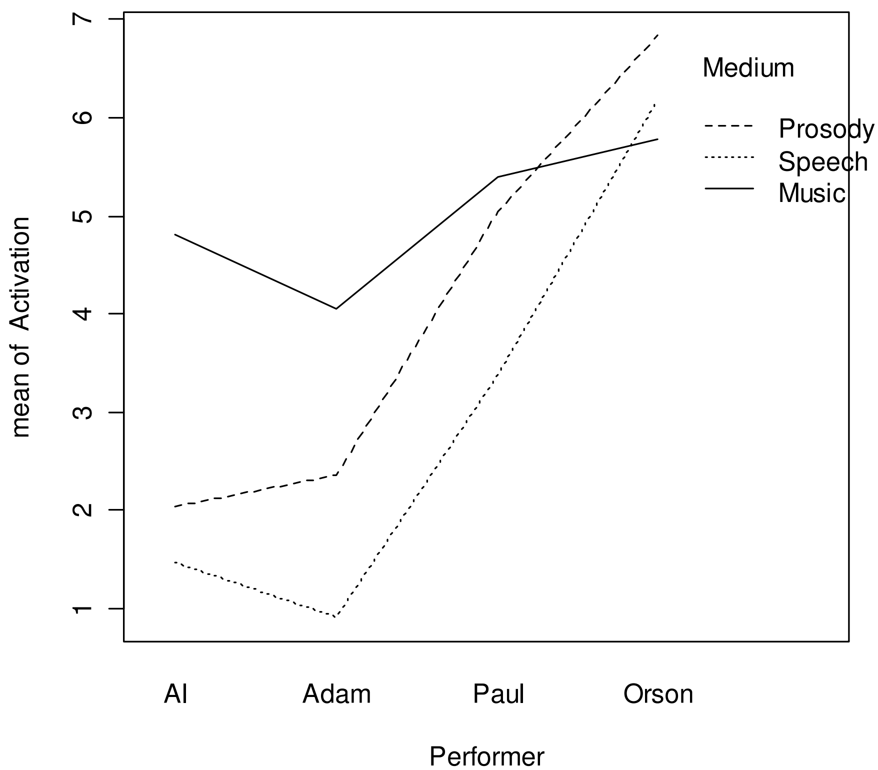

Figure 2. Main Effects of Performer and Medium on Activation

Once again, activation measures are inverted, so Adam has the highest mean activation ratings and Orson has the lowest for all three media. The patterns for activation generally stay the same relative to one another, demonstrating the performers' effects in spite of medium's absolute changes in activation value.

A second two-way repeated-measures analysis of variance showed main effects of performer and medium on participants' ratings of efficiency as well, but, again, no interaction effects, indicating additive and independent effects.

Error: Within

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Medium	2	180.94	90.47	19.0065	1.313e-08 ***
Performer	3	50.15	16.72	3.5116	0.01538 *
Medium:Performer	6	16.47	2.75	0.5768	0.74885
Residuals	396	1884.98	4.76		

Figure 3 shows these relationships graphically.

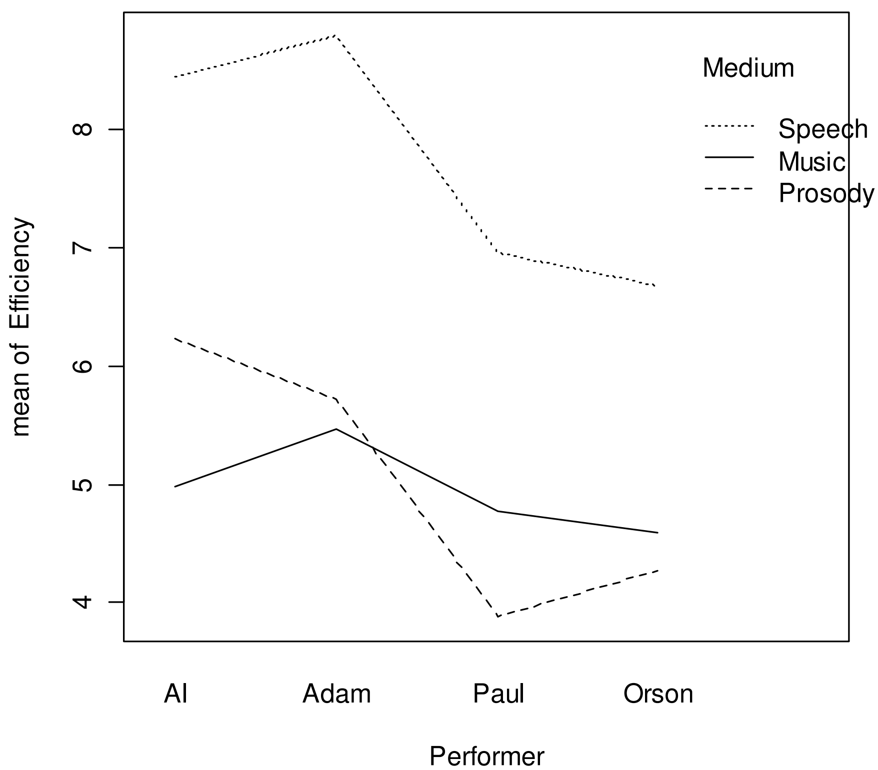

Figure 3. Main Effects of Performer and Medium on Efficiency

Like activation, efficiency patterns among the four performers generally remain the same relative to one another across the three media. In fact, Hierarchical Linear Modeling showed that both activation and efficiency could predict performer significantly:

Activation Fixed Effect	Standard Coefficient	Approx. Error	T-ratio	d.f.	P-value
For INTRCPT1, P0					
INTRCPT2, B00	2.500000	0.047451	52.686	34	0.000
For ACTIVATI slope, P1					
INTRCPT2, B10	0.195883	0.016691	11.736	418	0.000

Efficiency Fixed Effect	Standard Coefficient	Approx. Error	T-ratio	d.f.	P-value
For INTRCPT1, P0					
INTRCPT2, B00	2.500000	0.053142	47.044	34	0.000
For EFFICIEN slope, P1					
INTRCPT2, B10	-0.101295	0.020276	-4.996	418	0.000

The relative order of ratings across the four performers was similar for activation and efficiency, and differed at the same points in the prosody medium. This suggested a relationship between efficiency and activation, which was subsequently found. A correlation test for activation and efficiency showed an artificially negative relationship ($r = -0.4849$, $p < 2.2^{-16}$), which actually indicates that the greater the activation ratings for a sound file, the more efficient participants perceived it to be. Hierarchical Linear Modeling of this relationship confirmed its strength, and showed that activation predicts efficiency across all media and performers (with no effects of level 2 variables), and efficiency predicts activation across all media and performers (with a gender trend).

Fixed Effect	Coefficient	Standard Error	Approx. T-ratio	d.f.	P-value
For INTRCPT1, P0					
INTRCPT2, B00	5.895290	0.165510	35.619	34	0.000
For ACTIVATI slope, P1					
INTRCPT2, B10	-0.451497	0.041858	-10.786	29	0.000
GENDER, B11	-0.019564	0.086288	-0.227	29	0.822
PRIVATE0, B12	0.003575	0.012422	0.288	29	0.775
PROSODY, B13	-0.046185	0.045786	-1.009	29	0.322
IMAGININ, B14	-0.043420	0.057653	-0.753	29	0.457
DISTRACT, B15	-0.035083	0.088636	-0.396	29	0.695

The outcome variable is EFFICIEN

Fixed Effect	Coefficient	Standard Error	Approx. T-ratio	d.f.	P-value
For INTRCPT1, P0					
INTRCPT2, B00	4.130631	0.128700	32.095	34	0.000
For EFFICIEN slope, P5					
INTRCPT2, B50	-0.383227	0.072697	-5.272	29	0.000
GENDER, B51	0.297760	0.148639	2.003	29	0.054
PRIVATE0, B52	-0.004134	0.021755	-0.190	29	0.851
PROSODY, B53	-0.190153	0.086985	-2.186	29	0.037
IMAGININ, B54	0.015954	0.099072	0.161	29	0.874
DISTRACT, B55	-0.133857	0.158421	-0.845	29	0.405

The outcome variable is ACTIVATI

A final rating that was influenced by performer was pleasantness. In a two-way repeated measures ANOVA testing for effects of performer and medium, a robust main effect of medium was found, but no effect of performer. However, there was a significant interaction effect between medium and performer for pleasantness.

Error: Within

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Medium	2	262.08	131.04	38.6661	4.579e-16 ***
Performer	3	5.75	1.92	0.5651	0.63830
Medium:Performer	6	43.35	7.23	2.1321	0.04888 *
Residuals	396	1342.03	3.39		

This relationship shows that the interaction effect of performer on pleasantness depends on the effect of medium as well. Figure 4 is a graph that demonstrates how performer affects pleasantness ratings differently for different media.

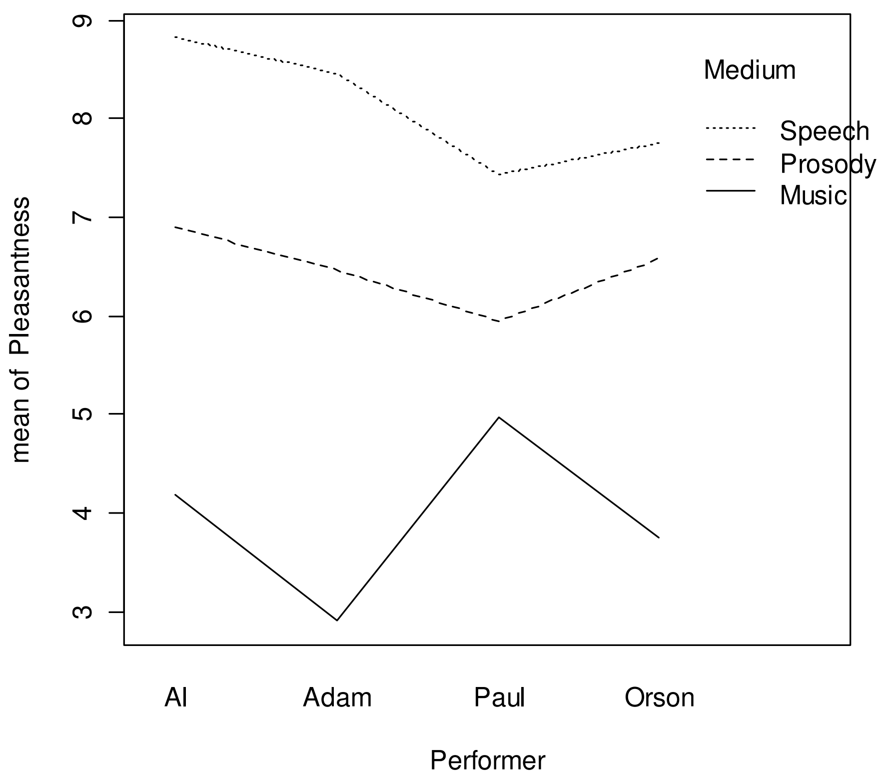

Figure 4. Main effect of Medium and Interaction Effect of Medium and Performer on Pleasantness

The relationships between performers seem to be similar for prosody and speech, but music is treated differently. Here is an instance of primary hypothesis 1 being disproven, where music and language seem to be beholden to different standards for ratings of pleasantness.

Primary Hypothesis 2

The second primary hypothesis, which holds that psychological ratings and physiological responses would be strongest for the speech stimuli, accounted for the possibility that though music and language might share similar qualities of emotional elicitation, they might elicit emotions to different degrees, especially speech in which words have the ability to clarify the precise emotion being expressed. This effect of medium is strong and has been shown in the three psychological ratings already discussed. For activation, efficiency and pleasantness, the strongest ratings are shown for the speech stimuli (pleasantness and activation are inverted) followed by prosody, then music stimuli. For pleasantness, the strongest valence was negative, which logically follows a monologue which is high in anger.

Liking is another rating in which this effect was found. Table 3 shows the basic numerical statistics for liking grouped by media.

	mean	sd	0%	25%	50%	75%	100%	n
speech	2.026857	4.182964	-6.09	-0.4525	2.21	5.2025	10.00	140
prosody	-2.565357	3.972193	-10.00	-5.1200	-2.49	0.0000	6.91	140
music	-0.061000	3.943829	-10.00	-2.2825	0.00	2.4125	9.18	140

Table 3. Liking

The two-way repeated-measures ANOVA for effects of medium and performer on liking showed only a main effect of medium.

Error: Within

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Medium	2	432.8	216.4	13.2158	2.779e-06 ***
Performer	3	39.8	13.3	0.8101	0.4888
Medium:Performer	6	34.3	5.7	0.3489	0.9104
Residuals	396	6484.6	16.4		

However, Table 3 and Figure 5 show that this effect of medium is acting differently on liking than on the other ratings. Here, though speech is still liked best, music is more preferred than prosody.

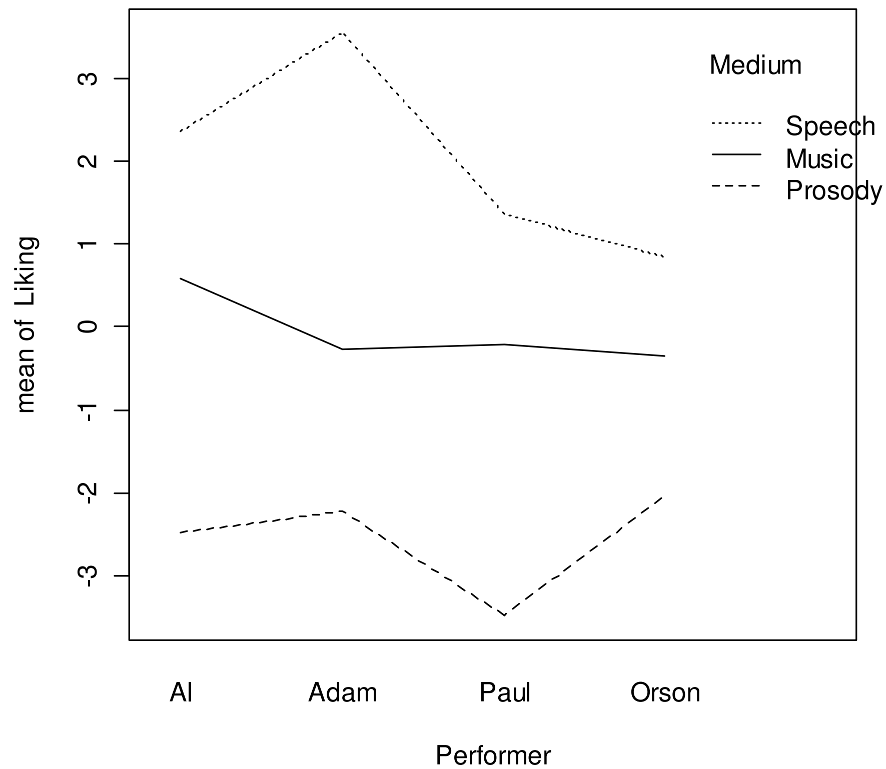

Figure 5. Main effect of medium on liking ratings.

Other significant relationships amongst psychological and physiological responses

None of the physiological measures showed any relationship with either the medium or performer factors and only limited relationships with the psychological ratings, but they were significantly related to one another (see Table 4).

	EMG	GSR	Heart.rate	Temperature
EMG	1.0000000	0.5227***	0.2395***	0.4996***
GSR	0.5227***	1.0000000	0.3231***	0.6304***
Heart.rate	0.2395***	0.3231***	1.0000000	0.4724***
Temperature	0.4996***	0.6304***	0.4724***	1.0000000

*Table 4. Correlation matrix of physiological responses. (*** = <0.000)*

Other significant relationships were found amongst the psychological ratings. Different HLM models show strong relationships between all four psychological ratings (Efficiency, Activation, Liking, Pleasantness). Some models of these relationships are provided. For Activation, Efficiency is the strongest predictor, but in models isolating Pleasantness and Liking, both variables can predict activation as well.

Fixed Effect	Coefficient	Standard Error	T-ratio	Approx. d.f.	P-value
For INTRCPT1, P0					
INTRCPT2, B00	4.021490	0.137794	29.185	34	0.000
For PLEASANT slope, P3					
INTRCPT2, B30	-0.104422	0.058844	-1.775	414	0.076
For LIKING slope, P4					
INTRCPT2, B40	0.058763	0.031258	1.880	414	0.060
For EFFICIEN slope, P5					
INTRCPT2, B50	-0.396633	0.057807	-6.861	414	0.000

The outcome variable is ACTIVATION

Efficiency is predicted by each of the three other psychological variables very strongly, with no effects of level two variables.

Fixed Effect	Coefficient	Standard Error	T-ratio	Approx. d.f.	P-value
For INTRCPT1, P0					
INTRCPT2, B00	5.885825	0.135268	43.512	34	0.000
For ACTIVATI slope, P2					
INTRCPT2, B20	-0.302302	0.045992	-6.573	29	0.000
GENDER, B21	0.036144	0.097798	0.370	29	0.714
PRIVATE0, B22	-0.004489	0.014288	-0.314	29	0.755
PROSODY, B23	-0.027975	0.053779	-0.520	29	0.606
For PLEASANT slope, P3					
INTRCPT2, B30	0.148302	0.053423	2.776	29	0.010
GENDER, B31	-0.110691	0.109534	-1.011	29	0.321
PRIVATE0, B32	0.011784	0.016580	0.711	29	0.483
PROSODY, B33	0.003699	0.055506	0.067	29	0.948
For LIKING slope, P4					
INTRCPT2, B40	0.283289	0.031438	9.011	29	0.000
GENDER, B41	0.036240	0.060970	0.594	29	0.556
PRIVATE0, B42	-0.010782	0.008962	-1.203	29	0.239
PROSODY, B43	0.000802	0.034025	0.024	29	0.982

The outcome variable is EFFICIENCY

Liking, a measure of particular interest, is best predicted by Efficiency and Pleasantness in a model involving the three other psychological variables, and though Activation did predict Liking in a simple model, it loses significance under the effects of Efficiency and Pleasantness.

Fixed Effect	Coefficient	Standard Error	T-ratio	Approx. d.f.	P-value
For INTRCPT1, P0					
INTRCPT2, B00	-0.067796	0.284509	-0.238	34	0.813
For ACTIVATI slope, P2					
INTRCPT2, B20	0.096821	0.090354	1.072	29	0.293
GENDER, B21	-0.121591	0.184694	-0.658	29	0.515
PRIVATE0, B22	0.015828	0.026880	0.589	29	0.560
PROSODY, B23	-0.015089	0.095626	-0.158	29	0.876
For PLEASANT slope, P3					
INTRCPT2, B30	-0.516263	0.121520	-4.248	29	0.000
GENDER, B31	-0.141870	0.249408	-0.569	29	0.573
PRIVATE0, B32	-0.035779	0.037528	-0.953	29	0.349
PROSODY, B33	0.088486	0.128422	0.689	29	0.496
For EFFICIEN slope, P4					
INTRCPT2, B40	0.985263	0.097259	10.130	29	0.000
GENDER, B41	-0.291525	0.203034	-1.436	29	0.162
PRIVATE0, B42	-0.012508	0.030672	-0.408	29	0.686
PROSODY, B43	0.055372	0.117036	0.473	29	0.639

The outcome variable is LIKING

Finally, Pleasantness is well-predicted by all three of the other variables without any effects of Level 2 variables.

Fixed Effect	Coefficient	Standard Error	T-ratio	Approx. d.f.	P-value
For INTRCPT1, P0					
INTRCPT2, B00	6.111752	0.102624	59.555	34	0.000
For ACTIVATI slope, P2					
INTRCPT2, B20	-0.094126	0.044627	-2.109	29	0.043
GENDER, B21	-0.114287	0.094893	-1.204	29	0.239
PRIVATE0, B22	0.000504	0.013662	0.037	29	0.971
PROSODY, B23	-0.006204	0.049573	-0.125	29	0.902
For LIKING slope, P3					
INTRCPT2, B30	-0.151917	0.036618	-4.149	29	0.000
GENDER, B31	0.000054	0.076988	0.001	29	0.999
PRIVATE0, B32	-0.001452	0.011020	-0.132	29	0.897
PROSODY, B33	-0.039848	0.042947	-0.928	29	0.362

For EFFICIEN slope, P4

INTRCPT2, B40	0.176652	0.058544	3.017	29	0.006
GENDER, B41	-0.120078	0.125199	-0.959	29	0.346
PRIVATE0, B42	0.008458	0.018477	0.458	29	0.650
PROSODY, B43	0.084097	0.070503	1.193	29	0.243
IMAGININ, B44	-0.025697	0.081666	-0.315	29	0.755
DISTRACT, B45	0.122066	0.133986	0.911	29	0.370

The outcome variable is PLEASANT

Secondary Hypothesis 1

A great deal of literature predicts that gender and musical training are two person factors that affect how well participants perceive and interpret emotional cues in voice and in music. Independent Samples T-Tests of gender were run for these data and several effects were found. Activation means for men and women were significantly different ($t(338.894) = 2.7364$, $p = 0.006539$). As Figures 6 and 7 show, women rate activation higher for all of the performers and for all of the media (lower means = higher activation).

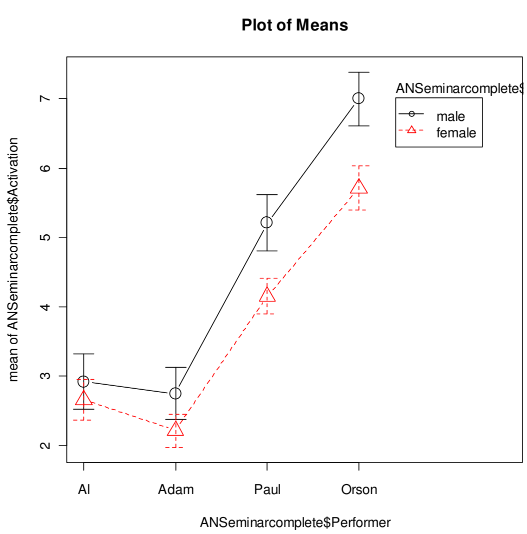

Figure 6. Activation means by gender and performer

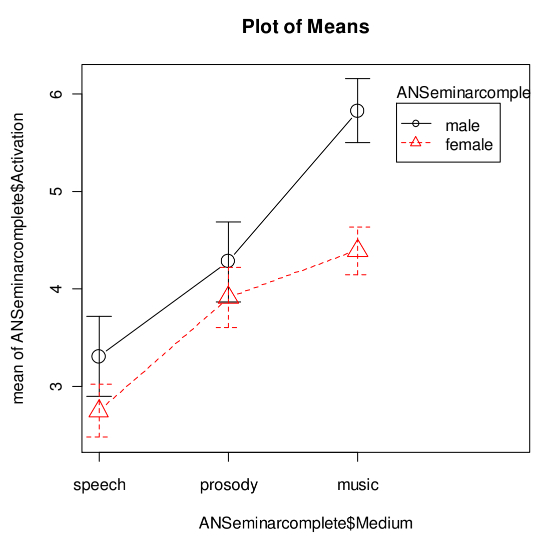

Figure 7. Activation means by gender and medium

In a similar way, heart rate means for men and women were significantly different ($t(265.917) = 3.5394$, $p = 0.0004732$). Figures 8 and 9 show that women have lower heart rate in response to all performers and all media.

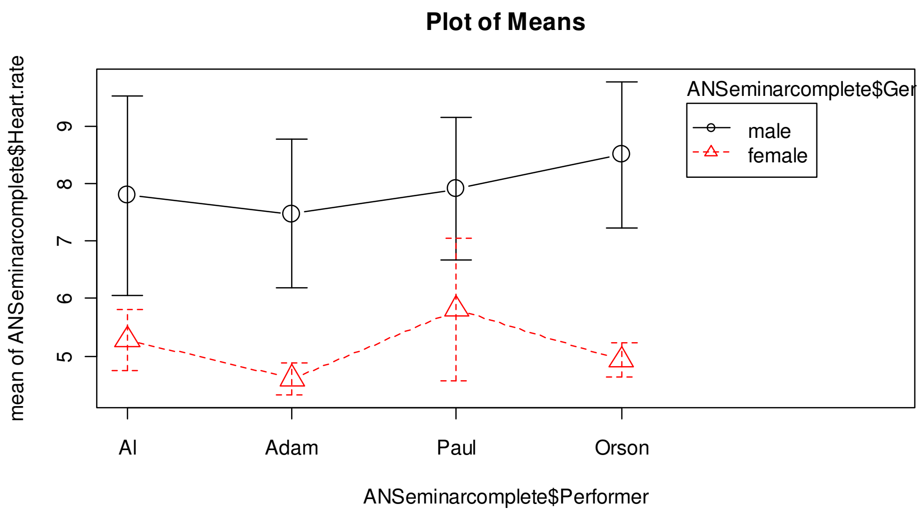

Figure 8. Heart rate means by gender and performer

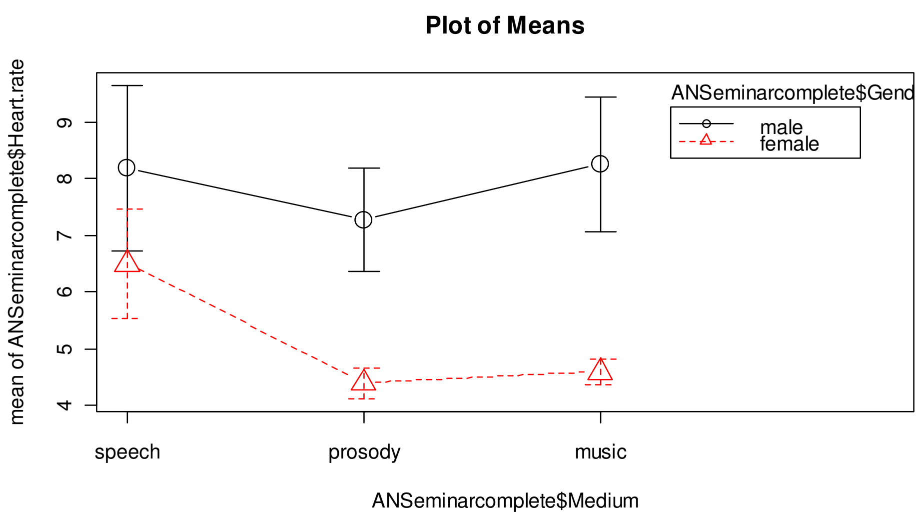

Figure 9. Heart rate means by gender and medium

Finally, this data produced confounding relationship between the two variables of interest, gender and music, as shown by the independent samples t-test results: (t(417.386) = -3.8306, p-value = 0.0001475). Women in this study had significantly greater musical training (as measured by private music lessons) that men.

In HLM analyses, Medium exhibited the influence of gender and music trends or significance in its relationship to Efficiency, Liking and Pleasantness, but neither of these Level

2 variables ever accounted for enough of the pattern that the significance of the Level 1 variable disappeared.

Secondary Hypothesis 2

The software package Praat was able to provide measures of pitch variables (i.e. minimum, maximum and average pitch, pitch range, and mean absolute slope – a measure of pitch variation) for all of the sound files. Correlational analysis in R showed that these pitch variables varied significantly with the psychological ratings across all media and performers in some cases.

Activation ($r = .2114$, $p<0.000$), Efficiency ($r = -.2089$, $p<0.000$) and Pleasantness ($r = -.3733$, $p< 0.000$) are correlated with minimum pitch of the piece so that when the minimum pitch is lower, activation is higher and efficiency and pleasantness are lower.

Efficiency ($r= .2631$, $p<0.000$) and Liking ($r =.3203$, $p<0.000$) are correlated with maximum pitch of the piece, so as the maximum pitch is higher, ratings of efficiency and liking are higher too.

Pleasantness ($r = -.4842$, $p<0.000$) correlates with average pitch in the piece, so that as the average pitch is higher, pleasantness is higher as well.

Finally, Efficiency ($r = .3211$, $p<0.000$) and Liking ($r = .3224$, $p<0.000$) are correlated with pitch range, so as range widens, so do ratings of efficiency and liking.

Discussion

Two primary hypotheses and two secondary hypotheses were developed for this experiment. Primary hypothesis 1 predicted that emotional responses, both psychological ratings and physiological responses, would follow the same patterns in all media derived from one of the four performers. Results indicate that this was not universally the case, but particular to certain psychological ratings. Activation and Efficiency both differ significantly by performer. These two measures exhibit roughly corresponding relationships so that in general, "Adam" pieces have the highest ratings of activation and the highest efficiency ratings, and "Orson" pieces, with the lowest activation scores are usually rated lowest on efficiency ratings too. The variation seems to come in the prosody pieces. Primary hypothesis 1, which hoped to find similar emotional responses across types of sound files, connects this prediction with the importance of the pitch (maintained across all media for one performer) in communicating emotional meaning in music, language, and–one might guess—all sound. This research might extrapolate that activation and efficiency are two conscious and controlled measures that have similar patterns or expectations for emotion across all sound.

Activation may have been significant in this study particularly because of the emotion with which the original monologue dealt. Shylock is angry, is feeling cheated and wronged, and is plotting revenge. Anger, as Scherer (1986) found, is associated with activation expectations for pitch variables, while pleasantness expectations for and responses to anger vary more by individual. That anger, carried in pitch variables, was transmuted through all of the performers and all of the pieces derived from those four original speech files. Therefore, activation and efficiency, a measure of how well expectations for acoustic expression of emotion are met, are two variables that we could foresee corresponding with the pitch choices of each performer.

Primary Hypothesis 2, that speech stimuli—having been clarified by spoken words—would elicit the strongest emotional responses, was largely confirmed by psychological ratings. Results corroborated the finding that content (verbal input) in matching its non-verbal elements of communication (and vice-versa) are the most positively received by listeners and add to their understanding of a message's relational meaning (Burgoon, Blair & Strom, 2008; Allan, 2006;

Mino, 1996; Kellaris & Kent, 1993; Markel, Bein, & Phillis, 1973). All four psychological ratings exhibited medium differentiation and speech had the greatest activation, lowest pleasantness (as expected for anger), greatest liking, and greatest efficiency ratings. Speech is the medium by which we are most accustomed to communicating and especially, clarifying emotional messages. Ratings for the prosody and music media also give us insight. For the same four ratings, prosody generally came next close to speech in terms of strong, anger-appropriate ratings. However, in the enigmatic measure of liking, after speech, music was the next most liked medium. Having also received the highest pleasantness scores, it is unclear whether the music stimuli at all conveyed the angry message contained in the speech and prosody. Arguably, music is generally listened to for pleasure, and though the music was not typical, it was probably hard to make the connection between it and the other two media. And perhaps the measures of liking and pleasantness are too individually-based—especially when experiencing anger—to produce general, significant results.

The secondary hypotheses refer to past literature about person characteristics and expectations for sound which may affect emotional experiences. Secondary hypothesis 1 predicted that women and more musically-trained individuals would have stronger emotional ratings and responses to all of the acoustical stimuli. Our results generally confirmed these past findings. T-tests showed significant differences between men and women in their activation ratings and heart rate responses. HLM analyses also indicated that gender affected all of the psychological ratings indirectly through their relationship with the “medium” factor. Unfortunately, a t-test also showed that musical-training, operationally-defined in this study as private lessons, was confounded with gender. Women participants had more musical-training than the men participants. Determining which of those had a greater effect on participants' ratings would be useful for future studies.

Secondary hypothesis 2 predicted that stimuli would be most preferred that best matched the literature findings for anger patterns (a high level, a wide range, and a large variability in pitch (Scherer, 1986; Juslin & Sloboda, 2001)) or had the greatest dynamism, which Mulac & Giles (1996) Addington (1971) and Black (1942) all found affected listener's preference for vocal delivery. Correlation data indicate these pitch characteristics associated with anger and dynamism had some effect on ratings. Measures of range (including minimum and maximum pitches) had diverse correlations with the psychological ratings. As expected, wider ranges and higher maximum pitches were associated with greater liking and ratings of efficiency. The higher maximum pitch corroborates the finding that angry sounds are higher in frequency, and a wider pitch range corresponds to greater pitch variance which is associated with both dynamism and anger literature.

What these pitch measures also revealed, however, is the pitch across a performer's speech, prosody and music files could not stay exactly the same. The process of getting a prosody file meant filtering out the top frequencies of a piece, so that alone caused changes in maximum pitch, pitch range and average pitch. There was also room for human error in the transcribing of prosograms into pure pitch music, because though prosograms produce readouts that correspond to semitones, the boundaries of each semitone is not perfectly clear and required human judgment. Nevertheless, finding some relationships between pitch variables, the performer factor and participants' ratings suggests that human error differences in a performer's stimuli did not completely erase the patterns of pitch that are important to emotion.

These control issues aside, the lab situation conceived for this experiment had the possibility of high internal validity, but not much external validity because the task is contrived and unlikely to be encountered in everyday life situations. Nonetheless, findings that pitch informs emotional communication could be influential in introductory encounters in which there is no preceding script for emotional relating between communicators and they therefore rely more heavily on non-verbal cues from the other person.

There are several future avenues for this research. The strong relationship that activation and efficiency had to performer in this experiment may be partially due to the importance of activation to anger. Performing similar experiments with stimuli charged with different emotions would be a good way to determine the specificity or generality of activation's importance.

Future studies might also benefit from musical stimuli that better corresponds to the speech stimuli, i.e. music with lyrics. Comparing vocal pitch movement with musical pitch movement, and music with words to voice with words could help to balance the experiment and clarify what influence each of those factors has on emotional understanding.

This experiment filled some holes in the literature especially in the nature of being a single continuous experiment examining musical and speech stimuli side-by-side. The effort to hold pitch constant across the stimuli of different media helped to clarify how pitch influences emotional communication. Results from this study largely confirm that there are expectations for how particular emotions, like anger, should sound. Those expectations appear to extend beyond just music or just speech. In music and in speech, we like dynamism, a characteristic that translates especially into activation, and pitch range and variance. In spite of a constant "contentual" message, we can make conscious evaluations of how efficient one speaker is from the next. And when asked, we can understand emotional content, not just in speech, but in music and just the sound of voice moving. We clue into pitch, and pitch gives us some emotional meaning.

Bibliography

Addington, D.W. (1971). The effect of vocal variation on ratings of source credibility. *Speech Monographs*, *38*(3), 242-247. Retrieved September 30, 2009 from EBSCO host.

Allan, D. (2006). Effects of popular music in advertising on attention and memory. *Journal of Advertising Research, 46*(4), 434-444.

Alter, K. & Knosche, T.R., (2003). Electrophysiological markers for phrasing in speech and music. (Abstract, Experimental Psychology Conference, 2003). *Australian Journal of Psychology*, supplement.

Barrett, L.F. (2006). Valence as a basic building block of emotional life. *Journal of Research in Personality, 40*, 35-55.

Barrett, L.F. (2005). Feeling is perceiving: Core affect and conceptualization in the experience of emotion. In L.F. Barrett, P.M. Niedenthal, & P. Winkielman (Eds.), *Emotions: Conscious and Unconscious* (pp. 255-284). New York: Guilford.

Barrett, L.F. (2004). Feelings or words? Understanding the content in self-report ratings of emotional experience. *Journal of Personality and Social Psychology, 87*, 266-281.

Barrett, L.F., & Bar, M. (2009). See it with feeling: Affective predictions in the human brain. *Philosophical Transactions of the Royal Society B: Biological Sciences, 394,* 1325-1334.

Barrett, L.F., & Bliss-Moreau, E. (2009). She's emotional. He's having a bad day: Attributional explanations for emotion stereotypes. *Emotion, 9,* 649-658.

Barrett, L.F., Bliss-Moreau, E., Quigley, K., & Aronson, K.R. (2004). Arousal focus and interoceptive sensitivity. *Journal of Personality and Social Psychology, 87,* 684-697.

Barrett, L.F., Lane, R., Sechrest, L., & Schwartz, G. (2000). Sex differences in emotional awareness. *Personality and Social Psychology Bulletin, 26*, 1027-1035.

Barrett, L.F., Lindquist, K., & Gendron, M. (2007). Language as a context for emotion perception. *Trends in Cognitive Sciences, 11*, 327-332.

Barrett, L.F., & Russell, J.A. (1999). Structure of current affect. *Current Directions in Psychological Science, 8,* 10-14.

Besson, M., Magne, C., & Schon, D. (2002). Emotional prosody: Sex differences in sensitivity to speech melody. *Trends in Cognitive Sciences, 6*(10), 405-407.

Black, J.W. (1942). A study of voice merit. *The Quarterly Journal of Speech, 28*(1), 67-74. Retrieved October 4, 2009, from EBSCO host.

Boersma, P., & Weenink, D. (2001). Praat, a system for doing phonetics by computer. Glot International.

Bruner, G.C. (1990). Music, mood, and marketing. *The Journal of Marketing, 54*(4), 94-104.

Burgoon, J.K., Blair, J.P., & Strom, R.E. (2008). Cognitive biases and nonverbal cue availability in detecting deception. *Human Communication Research, 34*(4), 572-599.

Crozier, W.R. (1997). Music and social influence. In D.J. Hargreaves & A.C. North (Eds.), *The social psychology of music*. (pp. 67-83). Oxford: Oxford University Press.

Curtis, M.E. & Bharucha, J.J. (in press). The minor third communicates sadness in speech, mirroring its use in music. *Emotion.*

Dellaert, F., Polzin, T., & Waibel, A. (1996). Recognizing emotion in speech. Presented at the Fourth International Conference on Spoken Language.

Duncan, S., & Barrett, L.F. (2007). Affect as a form of cognition: A neurobiological analysis. *Cognition and Emotion, 21,* 1184-1211.

Dutton, D.G. & Aron, A.P. (1974). Some evidence for heightened sexual attraction under conditions of high anxiety. *Journal of Personality and Social Psychology, 30*(4), 510-517.

Ferdinand, P. (2009). How music fine-tunes the brain.

Fortenbaugh, W.W. (1986). Aristotle's platonic attitude toward delivery. *Philosophy and Rhetoric, 19*(4), 242-254.

Fredrickson, B.L. (2000). Cultivating positive emotions to optimize health and well-being. *Prevention and Treatment, 3.* Retrieved from <http://journals.apa.org/prevention/volume3/pre0030001a.html>.

Frick, R.W. (1985) Communicating emotion: The role of prosodic features. *Psychological Bulletin,* 97, 412-29.

Geist, K., McCarthy, J., Rodgers-Smith, A., & Porter, J. (2008). Integrating music therapy services and speech-language therapy services for children with severe communication impairments: A co-treatment model. *Journal of Instructional Psychology, 35*(4), 311-316.

Herman, D. (2006). Prosodic foundations of language in-use. *American Speech, 81*(1), 94-99. Retrieved October 4, 2009, from EBSCO host.

<http://www.finalemusic.com>

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

Jennings, P., McGinnis, D., Lovejoy, S., & Stirling, J. (2000). Valence and arousal ratings for velten mood induction statements. *Motivation and Emotion, 24*, 285–297.

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

Kellaris, J.J., & Kent, R.J. (1993). An exploratory investigation of responses elicited by music varying in tempo, tonality and texture. *Journal of Consumer Psychology, 2*(4), 381-401.

Kraus, N., Skoe, E., Parbery-Clark, A., & Ashley, R. (2009) Experience-induced malleability in neural encoding of pitch, timbre & timing: implications for language and music. *Annals of the New York Academy of Sciences, 1169*, 543-557.

Lazarus, R.S. (1991). Cognition and motivation in emotion. *American Psychologist, 46*(4). 352-367.

Lee, K.M., Skoe, E., Kraus, N., & Ashley, R. (2009) Selective subcortical enhancement of musical intervals in musicians. *The Journal of Neuroscience, 29*(18), 5832–5840.

Lerdahl, F., & Jackendoff, R. (1983). *A generative theory of tonal music.* Cambridge, MA: MIT Press.

Magee, W.L., Brumfitt, S.M., Freeman, M., & Davidson, J.W. (2006). The role of music therapy in an interdisciplinary approach to address functional communication in complex neuro-communication disorders: A case report. *Disability and Rehabilitation, 28*(19), 1221-1229.

Markel, N.N., Bein, M.F., & Phillis, J.A. (1973). The relationship between words and tone-of-voice. *Language and Speech, 16*(1), 15-21.

Mauss, I.B. & Robinson, M.D. (2009). Measures of emotion: A review. *Cognition & Emotion, 23*(2), 209-237.

Mertens, P. (2005). The Prosogram.

MettingVanRijn, A. C., Kuiper, A. P., Dankers, T. E., and Grimbergen, C. A. (1996). Proceedings from the 18th Annual International Conference of the IEEE Engineering in

Medicine and Biology Society '96, *Low-cost active electrode improves the resolution in biopotential recordings*. Amsterdam, The Netherlands, Track 1.2.3-3.

Mino, M. (1996). The relative effects of content and vocal delivery during a simulated employment interview. *Communication Research Reports, 13*(2), p225-238.

Mulac, A., & Giles, H. (1996). 'You're only as old as you sound': Perceived vocal age and social meanings. *Health Communication, 8*(3), 199-215.

Mullennix, J.W., Bihon, T., Bricklemyer, J., Gaston, J., & Keener, J.M. (2002). Effects of variation in emotional tone of voice on speech perception. *Language and Speech, 45*(3), 255-283.

Musacchia, G., Sams, M., Skoe, E., & Kraus, N. (2007). Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. *Proceedings of the National Academy of Sciences, 104*(40), 15894-15898.

Musacchia, G., Strait, D., & Kraus, N. (2008) Relationships between behavior, brainstem and cortical encoding of seen and heard speech in musicians and non-musicians. *Hearing Research, 241*, 34–42.

Nolen-Hoeksema, S., Fredrickson, B.L., Loftus, G.R., & Wagenaar, W.A. (2009). *Atkinson & Hilgard's introduction to psychology.* Hong Kong: Cengage Learning.

O'Neill, S.A. (1997). Gender and music. In D.J. Hargreaves & A.C. North (Eds.), *The social psychology of music*. (pp. 67-83). Oxford: Oxford University Press.

Oudeyer, P.Y. (2003). The production and recognition of emotions in speech: Features and algorithms. *International Journal of Human-Computer Studies, 59,* 157-183.

Palmer, C. (1992). The role of interpretive preferences in music performance. In M.R. Jones & S. Holleran (Eds.), *Cognitive bases of musical communication* (pp. 249-262). Washington D.C.: American Psychological Association.

Patel, A.D. (2008). *Music, Language, and the Brain.* Oxford: Oxford University Press.

Patel, A.D., Peretz, I., Tramo, M., & Labreque, R. (1998). Processing prosodic and musical patterns: A neuropsychological investigation. *Brain and Language, 61,* 123-144.

Patterson, R.D., & Johnsrude, I.S. (2008). Functional imaging of the auditory processing applied to speech sounds. *Philosophical Transactions of the Royal Society B: Biological Sciences, 363*(1493), 1023-1035.

Pearce, W.B. (1971). The effect of vocal cues on credibility and attitude change. *Western Speech,* Summer, 176-184. Retrieved September 30, 2009, from EBSCO host.

Petty, R.E. & Cacioppo, J.T. (1986). The elaboration likelihood model of persuasion. *Journal of Personality and Social Psychology, 51*(5), 1032-1043.

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

Rosenberg, E.L. (1998). Levels of analysis and the organization of affect. *Review of General Psychology, 2,* 247-270.

Ross, D., Choi, J., & Purves, D. (2007). Musical intervals in speech. *Proceedings of the National Academy of Sciences, 104*(23). 9852-9857.

Sacks, O. (2007). *Musicophilia: Tales of music and the brain.* Toronto: Random House.

Schachter, S., & Singer, J. (1962). Cognitive, social and physiological determinants of emotional state. *Psychological Review, 69*, 379-399.

Scherer, K.R. (1986). Vocal affect expression: A review and a model for future research. *Psychological Bulletin, 99*(2), 143-165.

Scherer, K.R., Banse, R., & Wallbott, H.G. (2001). Emotion inferences from vocal expression correlate across languages and cultures. *Journal of Cross-Cultural Psychology, 32*(1), 76-92.

Scherer, K.R., Ladd, D.R., & Silverman, K.E.A. (1984). Vocal cues to speaker affect: Testing two models. *Journal of the Acoustical Society of America, 76*(5), 1346-1356.

Shaffer, L.H. (1992). How to interpret music. In M.R. Jones & S. Holleran (Eds.), *Cognitive bases of musical communication* (pp. 33-50). Washington D.C.: American Psychological Association.

Sloboda, J.A. (1992). Empirical studies of emotional response to music. In M.R. Jones & S. Holleran (Eds.), *Cognitive bases of musical communication* (pp. 33-50). Washington D.C.: American Psychological Association.

Spackman, M.P., Fujiki, M., Brinton, B., Nelson, D., & Allen, J. (2005). The ability of children with language impairment to recognize emotion conveyed by facial expression and music. *Communication Disorders Quarterly, 26*(3), 131-143.

Steeneken, H.J.M., & Hansen, J.H.L. (1999). Speech under stress conditions: Overview of the effect on speech production and on system performance. Presented at the IEEE International Conference on Communications.

Stegemoller, E.L., Skoe, E., Nicol, T., Warrier, C.M., & Kraus, N. (2008). Music training and vocal production of speech and song. *Music Perception, 25*(5), 419-428.

Steinbeis, N. & Koelsch, S. (2007). Shared neural resources between music and language indicate semantic processing of musical tension-resolution patterns. *Cerebral Cortex,(18).* 1169-1178.

Strait, D.L., Kraus, N., Skoe, E., & Ashley, R. (2009). Musical experience and neural efficiency: effects of training on subcortical processing of vocal expressions of emotion. *European Journal of Neuroscience, 29*(3), 661-668.

Thompson, W.F., Schellenberg, E.G., & Husain, G. (2004). Decoding speech prosody: Do music lessons help? *Emotion, 4*(1), 46-64.

Ververidis, D., & Kotropoulos, C. (2006). Emotion speech recognition: Resources, features and methods. *Speech Communication, 48*(9), 1162-1181.

Watzlawick, P., Bavelas, J.B., & Jackson, D.D. (1967). *Pragmatics of human communication: A study of interactional patterns, pathologies, and paradoxes.* New York: Norton.

Wieczorkowska, A., Synak, P., Lewis, R., and Ras, Z. (2005) Extracting Emotions from Music Data. 15th International Symposium on Methodologies for Intelligent Systems ISMIS 2005, Saratoga Sprins, NY, USA

Winkielman, P. & Berridge, K. (2003). What is an unconscious emotion? The case for unconscious 'liking'. *Cognition and Emotion, 17, 181-211.*

Wolfe, J. & Powell, E. (2006). Gender and expressions of dissatisfaction: A study of complaining in mixed-gendered student work group. *Women and Language, 29*(2), 13-21.

Wong, P.C.M., Skoe, E., Russo, N.M., Dees, T., & Kraus, N. (2007) Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nature neuroscience, 10*(4), 420-422.

Zatorre, R.J. & Gandour, J.T. (2008). Neural specializations for speech and pitch: moving beyond the dichotomies. *Philosophical Transactions of the Royal Society: Biological Sciences, 363.* 1087-1104.

Zhu, R., & Meyers-Levy, J. (2005). Distinguishing between the meanings of music: When background music affects product perceptions. *Journal of Marketing Research, 42,* 333-345.

Appendix

Shylock's Monologue (*The Merchant of Venice*)

He hath disgraced me, and hindered me half a million, laughed at my losses, mocked at my gains, scorned my nation, thwarted my bargains, cooled my friends, heated mine enemies; and what's his reason? I am a Jew. Hath not a Jew eyes? Hath not a Jew hands, organs, dimensions, senses, affections, passions? Fed with the same food, hurt with the same weapons, subject to the same disease, healed by the same means, warmed and cooled by the same winter and summer, as a Christian is? If you prick us, do we not bleed? If you tickle us, do we not laugh? If you poison us, do we not die? And if you wrong us, shall we not revenge? If we are like you in the rest, we will resemble you in that. If a Jew wrong a Christian, what is his humility? Revenge. If a Christian wrong a Jew, what should his sufferance be by Christian example? Why, revenge. The villainy you teach me I will execute, and it shall go hard but I will better the instruction.