6-9-2006

Perception of Emotion from Facial Expression and Affective Prosody

Noelle Turini Santorelli

PERCEPTION OF EMOTION FROM FACIAL EXPRESSION AND AFFECTIVE PROSODY

by

NOËLLE TURINI SANTORELLI

Under the Direction of Diana L. Robins

ABSTRACT

Real-world perception of emotion results from the integration of multiple cues, most notably facial expression and affective prosody. The use of incongruent emotional stimuli presents an opportunity to study the interaction between sensory modalities. Thirty-seven participants were exposed to audio-visual stimuli (Robins & Schultz, 2004) including angry, fearful, happy, and neutral presentations. Eighty stimuli contain matching emotions and 240 contain incongruent emotional cues. Matching emotions elicited a significant number of correct responses for all four emotions. Sign tests indicated that for most incongruent conditions, participants demonstrated a bias towards the visual modality. Despite these findings, specific incongruent conditions did show evidence of blending. Future research should explore an evolutionary model of facial expression as a means for behavioral adaptation and the possibility of an “emotional McGurk effect” in particular combinations of emotions.

INDEX WORDS: Bimodal emotion perception, emotion processing, facial expressions, affective prosody

PERCEPTION OF EMOTION FROM FACIAL EXPRESSION AND AFFECTIVE PROSODY

by

NOËLLE TURINI SANTORELLI

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of Master of Arts

in the College of Arts and Sciences

Georgia State University

2006

PERCEPTION OF EMOTION FROM FACIAL EXPRESSION AND AFFECTIVE PROSODY

by

NOËLLE TURINI SANTORELLI

Major Professor: Diana L. Robins
Committee: Roger Bakeman
Erin B. McClure

Electronic Version Approved:

Office of Graduate Studies
College of Arts and Sciences
Georgia State University
May 2006

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Page

Introduction

Human beings must be able to understand the emotions of others in order to engage in successful social interaction. Knowing how others are feeling is fundamental to communication success, social well-being, and adjustment (Ambady & Gray, 2002). Perception of emotion in the face and voice has a central function in communication and normally proceeds effortlessly and accurately. The failure to recognize or identify emotional expressions can thus have wide-reaching and long-term detrimental effects upon social behavior, and may serve as a risk factor for maladjustment and later adverse outcomes (Izard, 1977).

A core component of many psychiatric illnesses is poor social functioning, which appears to be associated with impaired or inappropriate recognition and regulation of emotional behavior (Herba & Phillips, 2004). In fact, abnormalities in emotion expression recognition have been associated with psychiatric disorders in both adult (de Gelder, Vroomen, Annen, Masthoff, & Hodiamont, 2003; de Gelder, Vroomen, de Jong, Masthoff, Trompenaars, & Hodiamont, 2005) and child populations (Crick & Dodge, 1994; McClure & Nowicki, 2001). Walker (1981) examined emotion expression recognition in childhood psychiatric populations. Children with schizophrenia were less accurate than aggressive, anxious-depressed, and typical children at recognizing emotion expressions. McClure & Nowicki (2001) examined the relationship between social anxiety and children's ability to decode nonverbal emotional cues. Their results indicated that difficulty identifying the emotions conveyed in children's and adults' voices was associated with general social avoidance and distress. Deficits in emotion expression recognition also have been reported in physically abused children. Pollack & Kistler (2002) reported that physically abused children label ambiguous expressions as angry more frequently than nonabused children. Perhaps the most widely studied area in terms of developmental psychopathology and emotional deficits is that of autism. A rich literature examines the nature of social deficits seen in autism spectrum disorders (ASD), specifically impairments in emotion processing (e.g., Dawson, Meltzoff, Osterling, & Rinaldi, 1998; Hill, Berthoz, & Frith, 2004). Individuals with ASD have difficulty identifying facial expressions (Adolphs, Sears, & Piven, 2001; Celani, Battacchi, & Arcidiacono, 1999; Gepner, de Schonen, & Buttin, 1994; Hobson, Ouston, & Lee, 1988; MacDonald et al., 1989; Yirmiya, Sigman, Kasari, & Mundy, 1992) and affective prosody (Boucher, Lewis, & Collins, 2000; VanLancker, Cornelius, & Kreiman, 1989), compared to children who were matched on either chronological or mental age.

Development of emotional processing and recognition

Few studies have explored the development of emotion expression recognition throughout childhood and adolescence. Additionally, there has been no generally accepted theory of emotion processing, nor any generally accepted theoretical framework within which to understand the development of emotion processing and associated neural systems (McClure, 2000). Numerous and varied perspectives on emotion processing make it particularly difficult to study the development of emotion processing abilities (Herba & Philips, 2004). For example, perspectives on emotion processing include investigation of the physiological experiential, cognitive, behavioral/expressive, attitudinal, and regulatory components (Brody, 1985). These varied perspectives of emotion processing call for the use of varied methodologies and therefore make it particularly difficult to study the development of emotion processing abilities (Brody, 1985; Plutchik, 1984).

Despite the paucity of research exploring the development of emotion processing throughout childhood and adolescence, there is a large body of literature on emotion expression recognition in infancy, providing evidence of remarkable abilities at a very young age. Reviews of the literature on the development of emotion recognition in human infants agree that infants are able to discriminate between expressions of positive and negative emotion by about three months and discriminate among negative expressions by 6 or 7 months (e.g., Nelson, 1987). Infants not only recognize emotion expressions but infants also produce facial expressions for interest, pain, and disgust. By the time the infant is 2 to 3 months old, adult observers can also distinguish expressions of anger and sadness, with expressions of fear appearing by 6 or 7 months (Izard & Harris, 1995).

Furthermore, some studies implicate the importance of multiple factors (i.e., inclusion of vocal information, or use of dynamic faces) in infants' ability to recognize emotion expression (Izard, 1977). In a developmental study of the bimodal perception of emotions, infants were presented with faces combined with voices. Five- to seven-month-old infants looked longer at a face displaying an emotion congruent with the tone of voice than at a face displaying an emotion incongruent with the tone of voice (Walker & Grolnick, 1983). Despite the abundance of research on emotion processing in infants, it is difficult to make comparisons across development. Much of the literature has focused on the infant and preschool periods (see McClure, 2000 for a review of studies). Additionally, studies that have explored emotional development in childhood have tended to focus on narrow age ranges (Herba & Phillips, 2004). Therefore, little is known about the continued course of emotional development throughout different stages of the life cycle. Furthermore, methodological discrepancies (i.e., the use of different dependent measures across development) between studies make comparisons across findings and age groups very difficult. In fact, these methodological discrepancies have led researchers to question whether or not the same construct of emotion processing is being measured throughout the different stages of development (McClure, 2000). However, further subtleties in emotion processing across the developmental stages may be assessed by using more realistic and ecologically valid stimuli (Herba & Philips, 2004).

Emotion processing and its neuroanatomical substrates

In regard to the anatomical circuitry underlying social perception, studies of lesioned animals and brain-damaged individuals suggest that the medial temporal lobe, especially the amygdala, is critical (Adolphs, 2001; Aggleton, 1992; Baron-Cohen et al., 1999, 2000; Dawson, 1996; Emery, et al., 2001). Patients who have had their amygdala removed or have partial bilateral lesions of the amygdala show impaired ability to recognize and match certain emotions, identify eye gaze directions, imagine emotional expressions and interpret social signals from the face (e.g., Young, Hellawell, Van De Wal, & Johnson, 1996). These impairments are also present when the emotion is verbally expressed instead of visually expressed (Scott et al., 1997), and are greater for fear than other emotions (Adolphs, Tranel, Damasio, & Damasio, 1994). A number of neuroimaging studies have been published on the localization of emotion processing in the human brain. These studies have shown that not only subcortical areas (e.g., the amygdala), but also cortical areas (e.g., the prefrontal cortex, cingulate cortex, and temporal cortices), are crucial in emotion processing (Phan, Wager, Taylor, & Liberzon, 2002; Phillips, Drevets, Rauch, & Lane, 2003; Esslen, Pascual-Marqui, Hell, Kochi, & Lehmann, 2004).

One of the best-known methods for eliciting emotion processing is the presentation of facial expressions. In addition to brain areas important for emotion perception there are brain regions that are preferentially activated by faces. Functional imaging studies have shown preferential activation in face-responsive regions in the fusiform gyrus and superior temporal sulcus during facial expression viewing. Face-responsive regions in the fusiform gyrus and superior temporal sulcus likely evolved as part of a distributed neural system for processing faces; however, this neural system is not specific only to facial stimuli, but also non-face stimuli that make similar computational demands (Adolphs, 2001; Gauthier, Tarr, Anderson, 1999; Gauthier, Skudlarski, Gore & Anderson, 2000).

Despite the general agreement across researchers on the brain areas utilized for the recognition of facial expressions, most authors have frequently disagreed on the specific regions of the brain that are activated or deactivated during emotion processing in general. For instance, some reported activation of the amygdala during emotion processing (Blair, Morris, Frith, Perrett, & Dolan, 1999; Breiter et al., 1996; Morris et al., 1996; Phillips et al.,1997), whereas others did not replicate this finding (e.g., Damasio et al., 2000; Sprengelmeyer, Rausch, & Przuntek, 1998). Often, two studies, using the same methods and stimuli to investigate the same emotion, report different findings regarding areas of brain activity. For example, Damasio et al. (2000) and Kimbrell et al. (1999) both used positron emission tomography (PET) to explore anger. Additionally, both studies induced anger through the recall of life events. Damasio and colleagues reported that emotional recall engaged cortical and subcortical regions of the brain (specifically the insular cortex, secondary somatosensory cortex, anterior and posterior cingulate cortex, hypothalamus, and nuclei in the brainstem tegmentum) although the patterns of activation/deactivation varied with each emotion-feeling cycle. However, Kimbrell and colleagues reported that compared to neutral emotion induction, anger induction was uniquely associated with increased cerebral blood flow in the right temporal pole and thalamus. Similarly, Breiter and colleagues (1996) and Sprengelmeyer and colleagues (1998) both used functional magnetic resonance imaging (fMRI) and Ekman and Friesen's pictures of facial affect (1976) to investigate fear perception; Breiter and colleagues reported that the amygdala was preferentially activated in response to fearful versus neutral faces, but Sprengelmeyer and colleagues reported that fearful expressions (as compared to neutral expressions) resulted in activation in the right fusiform gyrus and the left dorsolateral frontal cortex. Thus, even minor differences in experimental design, recording strategies, and analysis procedures might lead to different results (Esslen et al., 2004).

One hypothesis suggested to explain the discrepancies regarding areas of brain activity is that because emotion processing is essentially an evaluation strategy, it may be reasonable that several brain structures would have to integrate information to evaluate a given situation (Esslen et al., 2004). In fact, Esslen suggests that there are no cortical “emotion centers,” but rather a set of cortical regions that become activated at different times during emotion processing. Herba and Phillips (2004) suggest moving toward a systems-based approach that acknowledges the activation of and interaction among multiple brain areas when socially relevant stimuli are processed. Along these lines, functional connectivity suggests that higher brain functions, such as emotion processing, are results of interactions between functionally specialized brain regions. A functional connectivity model suggests that it is likely that the functions of even highly segregated brain regions are coordinated during perception and action (Roelfsema, Engel, Konig, & Singer, 1997). Moreover, socially relevant stimuli encountered in the natural environment are usually experienced through multiple sensory channels (e.g. sight, hearing, touch, smell, etc.). For example, we can *hear* and *see* a laughing face, or *see* a burning fire and *smell* smoke (de Gelder, Vroomen, & Pourtois, 2004). Thus, the idea of functional connectivity between multiple brain areas in response to socially relevant stimuli appears to be consistent with the fact that individuals experience social cues through multiple sensory modalities.

There are as of yet only a few general theoretical suggestions in the literature concerning the neuroanatomical correlates of multisensory integration (e.g., Damasio, 1989; Ettlinger & Wilson, 1990; Mesulam, 1998). Research suggests that the integration of sensory input from multiple modalities may occur in specialized areas of the brain, including the parietal lobe, superior temporal sulcus, and insula (Calvert, 2001; Calvert, Hansen, Iversen, & Brammer, 2001; Olson, Gatenby, & Gore, 2002) or in general sensory regions such as the claustrum (Olson et al., 2002), a subcortical structure lateral to the basal ganglia; superior colliculi (Bushara, Grafman, & Hallett, 2001; Calvert et al., 2001), a midbrain structure involved in the primitive visual system; and the amygdala (Amaral, Bauman, & Schumann, 2003), a limbic structure essential for processing emotion and detecting salience. The use of bimodal stimuli in future experiments may greatly enhance the ecological validity of emotion processing studies, and thus provide a more comprehensive understanding of the neural pathways involved in emotion processing (de Gelder & Vroomen, 2000; de Gelder & Bertelson, 2003).

Before turning to multimodal emotion perception and the issues of common processing resources it is important to review the current literature on the recognition of face and voice expressions.

Emotion Inferences from Facial Expressions

Although individuals are able to attend to both auditory and visual information, most adults show perceptual biases towards visual stimuli (Robinson & Sloutsky, 2004). In fact, Robinson and Sloutsky (2004) examined the processing of auditory and visual information and its changes in the course of development and found that while infants demonstrated an auditory preference, 4-year-olds switched between auditory and visual preference, and adults demonstrated a visual preference. Many studies on emotion recognition have used faces as stimuli to the point of suggesting that the face is the most telling bearer of an individual’s emotional state (deGelder, 2000). Research on the universality of facial emotions is usually traced back to Darwin’s (1872) views on the function of facial expressions. The Darwinian tradition in the study of emotion is associated most often with the research of Ekman and Izard

(Cornelius, 1996). Others argue against a primarily evolutionary model of facial expressions. They contend that facial expressions are fundamentally social tools for communicative purposes (Fridlund, 1992, 1994). Thus, as conceptualized from a social constructivist model, facial expressions are reflective of cultural differences, with little overlap in expressions across populations (see Averill, 1980). Although a detailed explanation of this debate is outside the scope of this paper, it is important to note the emphasis that has been placed on facial expressions in the exploration of emotional processing.

Emotion Inferences from Affective Prosody

While the perception of emotion from facial expressions has been widely studied, less is known about the ability to infer emotion from vocal cues or affective prosody. The accurate recognition of emotion from standardized voice samples using actor portrayals lies near 60%, which is about five times higher than what would be expected by chance (see Johnstone & Scherer, 2000). Several authors have suggested that there is a systematic correlation between emotions and acoustic parameters (Darwin, 1872; Ekman, 1992). Although pitch is generally thought of as the best indicator of affective prosody, duration and intensity have also been shown to play a role (Williams & Stevens, 1972; Murray & Arnott, 1993). Research has demonstrated the ability of individuals to recognize basic emotions expressed in both spoken sentences and isolated words (Pollack, Rubenstein, & Horowitz, 1960; Johnson, Emde, Scherer, & Klinnert, 1986; Scott et al., 1997). Although individuals tend to associate particular acoustic cues with discrete emotions, such vocal patterns in isolation are not always reliable indices of emotion (Scherer, Banse, Wallbott, & Goldbeck, 1991; deGelder, 2000). In fact, vocal emotions are not always easy to distinguish and may be marked by individual differences in the way speakers use acoustic parameters to express emotions (Lieberman & Michaels, 1962; Scherer, 1979).

Multisensory integration

Information about emotional states can be communicated through a variety of nonverbal behaviors, such as facial expressions, gestures, body postures, and tone of voice (Rothman & Nowicki, 2004). Individuals rarely experience sensory input from one modality in isolation. It follows, then, that perception of emotion in real world situations results from the integration of multiple cues, most notably facial expression and affective prosody (tone of voice). Most studies to date have explored static facial expressions of emotions; fewer studies have explored affective prosody. The bimodal perception of emotions (i.e., a situation in which the face and voice are presented together) presents a relatively less explored topic. Experimental paradigms incorporating audio-visual (AV) stimuli may lead to better understanding of emotion processing than has been possible in previous research utilizing static photographs. Thus, the use of AV stimuli would call for integration of emotional information from multiple modalities and subsequently offer a more ecologically valid approach to understanding emotion processing.

Given that the existing literature on bimodal emotion perception is sparse, it is helpful to examine a broader range of literature on the subject of integration. Numerous studies have shown the large impact that visual information can have on speech perception. The McGurk effect, as the phenomenon is now generally called, offers a particularly striking example of visual influence on speech perception. The McGurk effect demonstrates that speech information from the voice and concurrent presentation of incompatible speech information from the face lead to illusory percepts (McGurk & MacDonald, 1976). For example, when a spoken syllable like /ba/ is dubbed onto the visual presentation of a face articulating an incompatible syllable (i.e., /ga/) participants reported hearing a compatible syllable (i.e., /da/). When an auditory syllable is dubbed onto an incongruous visual syllable, the resulting percept is not either of the components, but an integration of information from both modalities. The participant’s identification judgments are influenced by both sources of information. When the participant heard the spoken syllable /ba/ and saw the facial presentation of the syllable /ga/, the participant extracted relevant information from each modality and combined it to form one unified concept: /da/. The percept that arises when auditory and visual signals are incongruent appears to be related to the quality of the information available from each modality (Jones & Callan, 2003). It is important to note that /ba/ is produced with closed lips, while /da/ and /ga/ are produced with open lips. The visual experience of a talker producing an open-lip sound seems to override the auditory experience of a closed-lip /ba/ syllable. Another form that will produce the illusion is /ma/ (auditory) + /ka/ (visual) = /na/. Again, /ma/ is produced with closed lips, while /ka/ and /na/ are produced with open lips (McGurk & MacDonald, 1976).

MacDonald & McGurk (1978) found that there were certain consonant combinations that elicited a greater effect than others. Consonants that use different formations of the mouth when spoken seem to have a greater influence on the McGurk effect than those that have the same mouth formations. Additionally, Green and Gerdeman (1995) found that when the auditory and visual stimuli contained different vowels, the effect of the McGurk illusion decreased significantly. Thus, the McGurk effect has been found to be stronger for certain combinations of syllables. Moreover, the McGurk effect is limited to similar sounds and some individuals do not experience a McGurk illusion regardless of the combination of syllables. Despite these limitations, studies have also examined a wide range of circumstances under which the McGurk effect occurs. Saldana and Rosenblum (1993) concluded that even when the facial images of the visual stimuli were blurred, the McGurk illusion was unaffected. It has even been found that prelinguistic infants exhibited the McGurk effect (Rosenblum, Schmuckler, & Johnson, 1997).

More recently, Massaro and Egan (1996) examined multisensory integration by focusing on the combination of auditory (tone of voice) and visual (facial expression) information in the course of emotion perception. Massaro and Egan presented their participants with a single word recorded by a speaker in one of three affective tones (happy, angry, or neutral) and showed them a computer-generated face displaying one of the same three emotions. The participant's task was to classify the emotion as happy or angry. The frequency of either response depended on the emotions expressed in both the face and the voice. The authors found a strong positive correlation between reaction time and a measure of the ambiguity of each input configuration regarding the target decision. The authors discussed their results in terms of a multiplicative model of feature integration known as the fuzzy logical model of perception (FLMP). In the FLMP model, inputs are processed separately and then integrated to form a single percept. The model assumes three basic stages of processing: (1) each source of continuous information is evaluated to ascertain the degree to which it matches various stored prototypes, (2) the sources are integrated according to a multiplicative formula to provide an overall degree to which they support each alternative, and finally, (3) a decision is made on the basis of relative goodness of match with each prototype (Massaro & Egan, 1996, p. 217). See Figure 1 for a schematic representation of these stages.

A contrasting model to the FLMP is the additive model of perception (AMP; Huber & Lenz, 1993). The AMP approach to emotion perception is equivalent to a categorical, or single-channel, model of perception. This model predicts that a participant will categorize an emotion from each modality and respond with the outcome from one of these categorizations (Massaro, 1987). Information is gathered by the separate modalities and is added together in the simplest manner, without regard to feedback across modalities (Bruno & Cutting, 1988). According to the

AMP, participants process all information sources separately, weight them, and then add the results to form the percept (Cutting, Bruno, Brady & Moore, 1992). In other words, this model allows one sensory modality to have more influence than the other. See Figure 2 for a schematic representation of this model.

Considering these two models, two general possibilities emerge regarding the integration of emotion processing: (1) either many sources are integrated in some manner (as the FLMP would suggest), or (2) emotional information is selected from one source (as the AMP would suggest; Cutting et al., 1992). Another way of viewing the difference between the two models of integration is in algebraic terms. Within information integration theory, models divide into two major classes. One consists of adding, subtracting, and averaging models that express the integration process as a weighted sum of information components; the other consists of multiplying and dividing models that use joint addition and multiplication rules. An integration model in its most general form could be written as:

$$R = f\,(c\,[w_1 * s_1,\; w_2 * s_2]),$$

where R is a response to be determined, s_1 and s_2 are two different sources of information, w_1 and w_2 are weights assigned to the sources, c is a combination rule, and f is the function that maps combined information to percept (Bruno & Cutting, 1988). The three stages of information integration are represented in the above equation: the evaluation process (equivalent to assessing weights, w), the integration process (which is reflected in the combination rule, c), and the classification process (which generates the response and is represented by f); (Bruno & Cutting, 1988).

In the FLMP, it is assumed that an individual multiplies the sources of information available. The simplest multiplication may be written as:

$$R = f\,(w_1 * s_1 * w_2 * s_2)$$

where *R* is a response to be determined, s_1 and s_2 are two different sources of information, w_1 and w_2 are weights assigned to the sources, and *f* is the function that maps combined information to percept (Cutting et al., 1992). In the AMP, it is assumed that the sources of information are added and the integration process is then expressed as a weighted sum of the information components (Bruno & Cutting, 1988). Symbolically, the simplest additive strategy is:

$$R = f\,(w_1 * s_1 + w_2 * s_2)$$

where *R* is a response to be determined, s_1 and s_2 are two different sources of information, w_1 and w_2 are weights assigned to the sources, and *f* is the function that maps combined information to percept (Cutting et al., 1992).

A study conducted by deGelder and Vroomen (2000), although similar to Massaro and Egan's (1996) work, included a number of important differences. deGelder and Vroomen were interested in determining if participants who were presented simultaneously with facial affect and affective prosody would combine the two sources of information to decide what emotion was presented, or base their response on only one modality. Participants were presented with a still photograph of a face on a screen while a voice was heard pronouncing a sentence in one of two tones, either sad or happy. The faces used in this study were taken from a morphed continuum extending between extreme sadness and happiness. The participants were asked to indicate, by pressing one of two keys, whether the person was happy or sad. When presented with both facial affect and prosodic affect, participants appeared to combine the two sources of information. The combination was seen in both the accuracy of responses and in the reaction times. The longest reaction times were obtained for incongruent stimuli (when the actor's facial expression was different from his or her tone of voice).

The authors also were interested in determining if the integration of cross-modal emotional stimuli is an automatic process. In Experiments 2 and 3 of the same study, deGelder and Vroomen instructed participants to base their response on the inputs of one of the modalities and ignore those of the other modality. Interestingly, despite instructions to focus on only one modality and ignore the other, participants were nevertheless influenced by the modality that was supposed to be ignored. The fact that biases occurred in spite of instructions to ignore the non-target stimuli may be supporting evidence for the automatic nature of bimodal integration and may provide support for the existence of what deGelder & Bertelson (2003) have termed an 'emotional McGurk effect.'

The Current Study

Despite the FLMP and the AMP models of emotion perception, to date there has been no generally accepted theory of multimodal emotion processing, nor any generally accepted theoretical framework with which to understand the development of emotion processing and associated neural systems (Herba & Phillips, 2004; McClure, 2000). A greater understanding of emotion recognition, a core social function, will provide valuable normative data. In the future, these data could be used to inform the identification of abnormal patterns of emotion recognition. Although many studies have examined emotion processing, most studies conducted in the laboratory are very different from the social encounters people experience in the real world. For example, Massaro & Egan (1996) and de Gelder and Vroomen (2000) obtained evidence of bimodal integration, although both studies utilized stimuli that lack ecological validity. It can be argued that human responses are relevant to the general issue of bimodal perception of emotions only if they reflect basic perceptual processes rather than specific strategies adopted to satisfy the demands of particular laboratory tasks. Thus, the goal of this study is to build upon previous studies (deGelder & Vroomen, 2000; Massaro & Egan, 1996) and further examine the integration of emotional cues from auditory and visual modalities. The current study attempts to generate more realistic emotion processing by using short movies instead of static photographs and disconnected audio recordings. Participants will watch short movies that contain both auditory and visual emotional cues. Participants will then decide what emotion the actor is portraying. Stimuli containing dynamic emotional information both in facial expression and tone of voice mimic the emotion processing abilities required in the real world, and may provide greater insight into social and emotional processing.

Hypotheses

The proposed research will be useful in determining how facial expressions and affective prosody are evaluated during the perception of emotion. More specifically, the proposed hypotheses are concerned with the combination of auditory and visual sources of information and its effect on the judgments made regarding the displayed emotions.

Hypothesis 1: When emotional cues are incongruent, more time will be required relative to congruent emotional cues before a sufficient degree of support accumulates and a response is emitted. Hence, hypothesis 1 is that mismatching movies will result in an increased reaction time or latency effect compared to matching movies.

Hypothesis 2: Hypothesis 2 will examine the prediction that when presented with matching emotional stimuli, participants will choose the correct emotion (the emotion that was presented by both the facial expression and tone of voice).

There are a number of possibilities regarding the effect of incongruent emotions on the participant's perception of those emotions.

Hypothesis 3: The first possibility is that participants will respond with a bias of facial expression when presented with a given incongruent emotional stimulus. Although participants will process both the auditory and visual information simultaneously, the final decision will only take into account the information supplied by the more salient modality, which is hypothesized to be the visual modality. It is hypothesized that the most salient modality will be the visual modality for a number of reasons. First, the finding that although individuals are able to attend to both auditory and visual information, adults tend to show perceptual biases toward visual stimuli (Robinson & Sloutsky, 2004). Second, the McGurk effect which demonstrates the ability of visual information to alter the percept of speech perception demonstrates the importance of visual information in cross-modal processing. Finally, the overall accuracy percentage for vocal emotion expressions is somewhat lower than that found in equivalent studies on the decoding of facial expressions (Scherer, 1999). Thus, participants may simply use the single most effective available source and disregard the others. For example:

Angry Face + Happy Voice → Angry (Evidence of facial bias)

Hypothesis 4: Along these lines, hypothesis 4 predicts that certain participants may consistently prefer one sensory modality over the other when presented with incongruent emotional stimuli. For example, certain participants may choose the emotion expressed in the face over the emotion expressed in the tone of voice as a general rule throughout the different trials. Similarly, certain participants may consistently choose a response other than that expressed by the face or voice. According to this hypothesis, the sample would be divided based on individual preferences or individual differences in emotion perception and therefore would not show an overall group bias towards a single sensory modality. In addition, there may be modality preference that is dependent upon group membership. For example, participants may prefer one sensory modality over the other depending on their sex.

Hypothesis 5: Certain incongruent conditions will elicit a response other than the emotions presented in the face or the voice. Thus, the participant will perceive a third, complex emotion rather than the emotion in the face or voice. In this case, we will see the integration of two modalities without a bias toward one sensory modality. It is important to note that this effect may not occur with all incongruent stimuli. Similar to the finding that the McGurk effect is stronger for certain combinations of syllables (e.g., /ba/ + /ga/ = /da/), there may be something inherent in certain combinations of emotions that elicit integration.

In order to test the specificity of these predictions, all analyses will be conducted after examining several potentially confounding variables. It has been suggested that individual differences in emotion processing will stem from multiple factors, including individual characteristics (e.g., cognitive ability; Bennett, Bendersky, & Lewis, 2005). Emotion processing is believed to reflect an individual's basic information-processing skills using emotional information as the relevant data (Salovey, Bedell, Detweiler, & Mayer, 2000). Thus, emotion processing and general cognitive ability may be closely related. Considering the relationship between general cognitive ability and emotion processing skills, basic perceptual abilities such as face recognition and sound perception, an estimate of general mental ability, and a mental health screening will be examined prior to all statistical analyses. Inclusion of these measures is important in order to rule out the possibility that participants' emotion processing skills are being influenced by confounding variables such as general mental ability and basic perceptual skills.

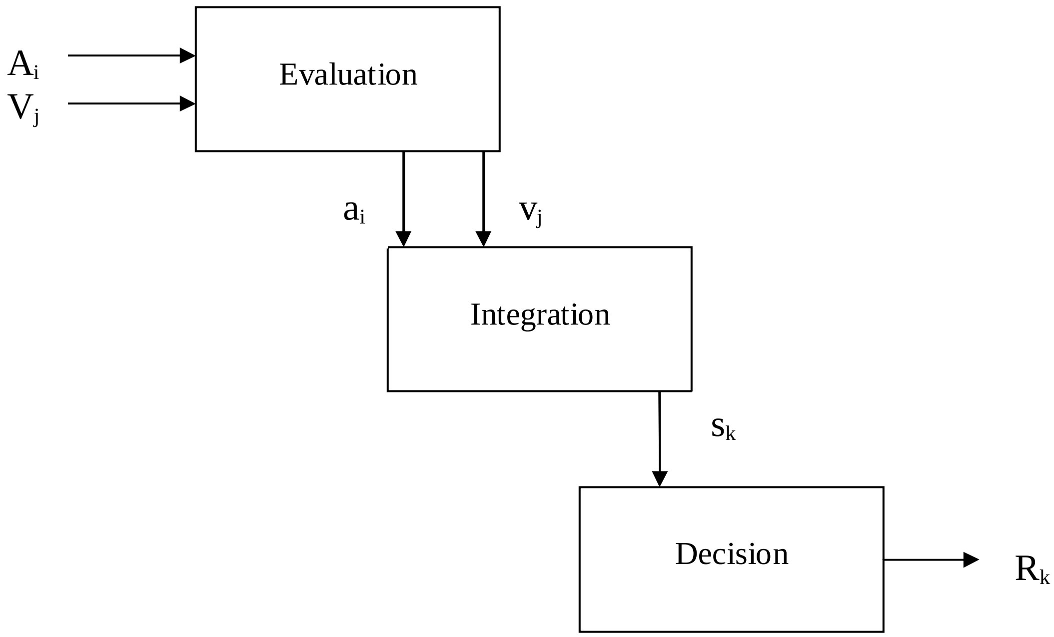

Figure 1 Schematic representation of the three stages involved in perceptual recognition according to the FLMP. The three stages are shown to proceed left to right in time to illustrate their necessarily successive but overlapping processing. The sources of information are represented by uppercase letters. Auditory information is represented by Ai, and visual information by Vj. The evaluation process transforms these sources of information into psychological (or fuzzy truth, Zadeh, 1965) values (indicated by lowercase letters ai and vj). These sources are then integrated to give an overall degree of support for a given alternative s_k. The decision operation maps this value into some response, R_k, such as a discrete decision or a rating (Massaro & Cohen, 2000, p. 314).

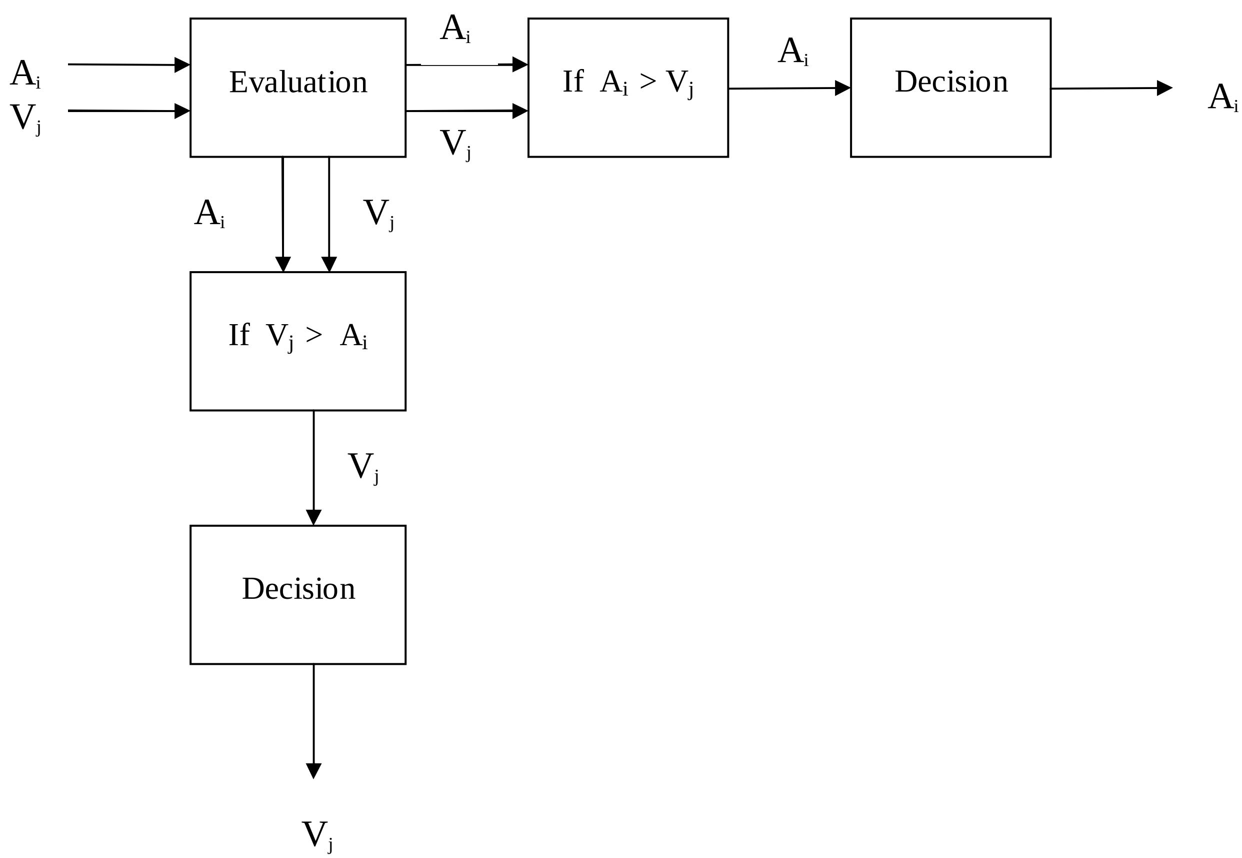

Figure 2 Schematic representation of the three stages involved in perceptual recognition according to the AMP. There are two possible decisions according to the AMP. The sources of information are represented by uppercase letters. Auditory information is represented by Ai and visual information by Vj. The evaluation process proceeds and determines whether Ai > Vj or Vj > Ai. The decision will be a response that reflects the greater of the two.

Method

Participants

Thirty-nine students from the undergraduate psychology subject pool at Georgia State University participated in the current study. The participants were registered in the subject pool as part of their required undergraduate psychology work. Participants ranged in age from 18-34 (M = 20.84). The sample consisted of 12 males and 27 females. Two participants met criteria for exclusion leaving a total sample size of 37. Participants in the final sample ranged in age from 18-34 years of age (M = 20.97). The final sample consisted of 10 males and 27 females.

Procedure

Data for the proposed study have been collected as part of a study developing normative data for a set of bimodal emotion stimuli. All participants provided written informed consent. Trained undergraduate and graduate students administered tasks individually to participants in three-hour testing sessions. Test sessions took place in a quiet room in the Georgia State University Psychology Clinic. Participants were offered the opportunity to take breaks as needed. Task order was randomly counterbalanced across participants to prevent order effects. No feedback was provided to the participants regarding performance on either the experimental or clinical measures. However, if a participant indicated severe psychopathology suggesting that he or she may be a danger to him/herself or others, that participant was contacted and appropriate referrals were made for mental health services. Participants were thoroughly debriefed when they had completed the experiment. Debriefing included an explanation of the experiment by the examiner and a handout was given to each participant that further explained the experiment and its research implications.

Measures

Wechsler Abbreviated Scale of Intelligence (WASI; Wechsler, 1999). Participants were administered the WASI to estimate overall cognitive ability. The WASI consists of four subtests (Vocabulary, Block Design, Similarities, and Matrix Reasoning) used to assess various aspects of intelligence. The WASI possesses adequate psychometric qualities. Reliability coefficients were developed for each of the subtests, as well as for the IQ scales. For adults, the reliability coefficients range from .84 to .98. Content validity was demonstrated by both content coverage and content relevance. In addition, the WASI is highly correlated with other ability and achievement measures (Psychological Corp., 1999). Participant's with IQ scores less than or equal to a score of 70 were eliminated from statistical analyses.

Diagnostic Analysis of Nonverbal Accuracy Scale - Second Edition (DANVA 2; Nowicki, 2004). Participants were administered all four subtests from the DANVA 2: Adult Facial Expressions, Adult Paralanguage, Child Facial Expressions, and Child Paralanguage. Each subtest consists of 24 trials (12 male, 12 female). Participants were asked to identify, using a forced-choice format, the facial expression depicted in the photograph (Adult and Child Facial Expressions) or to identify the emotion perceived from the actor's voice (Adult and Child Paralanguage). The participants were given the same four response choices (happy, sad, angry, and fearful) for each subtest. All subtests were computer-administered. These subtests were administered to assess the participant's ability to accurately decode nonverbal cues. For information on scale construction, see Nowicki & Duke, 1994 and Nowicki, 2004. Errors were summed across all four subtests. Data with a mean number of errors greater than or equal to two standard deviations below the mean were excluded from statistical analyses.

Adult Facial Expressions. The Adult Facial Expressions subtest of the DANVA-2 consists of 24 photographs of adults displaying one of four facial expressions: happy, sad, angry, and fearful (Nowicki & Carton, 1993). Construct validity has been reported by Nowicki (2004). During test administration, the photographs were presented, one at a time, each for a 2-second exposure period. Participants were asked to identify the facial expression depicted in the photograph.

Child Facial Expressions. The Child Facial Expressions subtest of the DANVA-2 consists of 24 photographs of children showing happy, sad, angry, and fearful faces. Construct validity information for Child Facial Expressions is available from 50 studies (Nowicki, 2004). During test administration, the photographs were presented, one at a time, each for a 2-second exposure period. Participants were asked to identify the facial expression in the photograph.

Adult Paralanguage. The Adult Paralanguage subtest of the DANVA-2 consists of male and female voices repeating a neutral sentence, “I am going out of the room now, but I’ll be back later” in happy, sad, angry, and fearful voices. Participants were asked to identify the emotion perceived from the actor’s voice. Participants were able to press a button to repeat each sentence as many times as necessary before giving a response.

The 24 items were selected from a pool of 133 recordings on the basis of high inter-rater agreement (70-80%) regarding the presented emotion. Baum & Nowicki (1998) presented data supporting the reliability and construct validity of this subtest. Results from eight studies showed coefficient alphas ranging from .71 in four-year-old subjects to .78 in college students, with a median coefficient alpha of .76. Test-retest reliability over six weeks was .83 in a sample of college students (n = 68, M = 19.4 years) (Baum & Nowicki, 1998).

Child Paralanguage. The Child Paralanguage subtest of the DANVA-2 consists of 24 trials in which several children repeat the sentence, “I am going out of the room now, but I’ll be back later,” in happy, sad, angry, and fearful voices. Participants were asked to identify the emotion perceived from the actor's voice. Participants were able to press a button to repeat each sentence as many times as necessary before giving a response. Construct validity support from more than 50 studies has been presented by Nowicki & Duke (1994), Nowicki (2004), and Rothman & Nowicki (2004).

Benton Facial Recognition Test (BFRT; Benton, 1994). The BFRT examines the ability to recognize faces without a memory component. Participants were exposed to three different matching conditions: matching of identical front views, matching of front-view with three-quarter views and matching of front-view under different lighting conditions. The test has 22 stimulus cards and calls for 54 separate matches. Six items call for one match to the sample photograph and 16 items call for three matches to the sample photograph. This test was administered in order to determine if participants have normal facial processing abilities. If a participant's performance on this test is below average it suggests that their performance on the experimental measure may be negatively impacted. Participants who do not score within the normal range were eliminated from statistical analyses.

Seashore Rhythm Test (Seashore, Lewis, & Saetveit, 1960). This test requires participants to discriminate between like and unlike pairs of rhythms. Derived from a subtest of the Seashore Measures of Musical Talent, the Rhythm Test requires participants to discriminate between 30 pairs of rhythmic beats as either different or the same. Classified as a measure of non-verbal auditory discrimination, the reaction time is particularly sensitive to the participant's ability to attend and concentrate. Test-retest differences are small (McCaffrey, Duff, & Westervelt, 2000) and internal reliabilities (split-half and odd-even) of .77 and .62 have been reported (McCaffrey et al., 2000). This test was administered to assure that each participant has normal ability for auditory discrimination. If a participant's performance on this test is below average, it suggests that their ability to accurately decode affective prosody on the experimental measure may be negatively impacted. Data with a mean number of errors greater than or equal to two standard deviations below the mean were excluded from statistical analyses.

Adult Self-Report Inventory – Fourth Edition (ASRI-4; Gadow, Sprafkin, & Weiss, 1999). The ASRI-4 is a symptom rating scale that can be used to screen for behavioral, emotional, and cognitive symptoms defined in the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV; American Psychological Association,1994). Participants indicate the frequency of occurrence of each symptom (i.e., never, sometimes, often, or very often). Individual symptoms are considered to be clinically significant if they occur "often" or "very often." The cutoff scores for the ASRI-4 are based on the number of symptoms that are rated as being clinically significant according to the DSM-IV. The items in the ASRI-4 are grouped according to diagnostic category. The ASRI-4 compares favorably with other scales and procedures; it is time-efficient and it offers an alternative to structured psychiatric interviews. Based on analyses of ratings for a normative sample, the internal consistencies of the ASRI-4 categories were generally high, with alphas above .70 except for the eating disorders and substance use categories and categories with as few as three items (i.e., Dissociative Disorder, Schizoid Personality Disorder). Occupation and education were minimally correlated with ASRI-4 Symptom Severity scores in the normative data sample (N = 900) (Gadow et al., 1999). The ASRI-4 was administered to assure that the presence of emotional or behavioral symptoms would not impact performance on the experimental measure. Participants with T-scores greater than or equal to 70 and meet criteria for the Symptom Count Cutoff score for the same diagnostic category were eliminated from statistical analyses.

Experimental Stimuli (Robins & Schultz, 2004). Novel AV stimuli were developed and validated at the Yale Child Study Center (Robins & Schultz, 2004), using professional actors (one male and one female) and a local video production studio. Actors delivered 10 sentences in four emotional tones (angry, fearful, happy, and neutral). See Figures 3 and 4. Sentences were emotionally ambiguous, (i.e., each was feasibly delivered in all four emotional tones). For example, "The door is open" can be angry if someone left the door open and rain flooded the hallway, fearful if the speaker fears an intruder opened the door, happy if the speaker is welcoming someone, or neutral as a simple statement of fact. The movies were separated into audio and video tracks and remixed, yielding 320 movie clips. Eighty contain audio and video tracks with matching emotions (i.e., Happy Face / Happy Voice), and the remaining 240 stimuli contained incongruent emotional cues (i.e., Happy Face / Angry voice). The 320 movie clips were derived by the following equation: (4 facial expressions) x (4 tones of voice) x (2 genders) x (10 emotionally ambiguous sentences) = 320. Matching movies were cross-spliced using two different recordings of the same emotion, to ensure that perception differences between the matching and mismatching movies were not due to method of development. Lip synchrony was maintained in all stimuli. Stimuli are approximately 1.5 - 2.5 seconds in duration. Sample stimuli can be found at http://www.gsu.edu/%7Ewwwpsy.faculty/robins.htm.

Stimuli were presented to participants using PsyScope (Cohen et al., 1994) on a G3 Macintosh iBook laptop computer running Macintosh OS 9 Classic operating system. Behavioral response and reaction time were collected by PsyScope. Participants were instructed to watch short movie clips played on the computer and identify the emotion they believed the actor was portraying. Participants were required to respond to each movie clip in a forced-choice format, by pressing a correspondingly labeled key on the computer keyboard. Participants were seated at approximately a 60-cm distance from the computer screen. Participants were instructed to choose the portrayed emotion based on their initial reaction and not to spend too much time on any one movie clip. See Figure 5 for a list of the emotions from which the participant was asked to select after viewing each movie clip. Six emotions were selected from Ekman & Friesen's (1976) basic emotions: happiness, surprise, fear, anger, disgust, and sadness. Each of these emotions can be reliably signaled by the face and have been extensively normed (e.g. Ekman & Friesen, 1986; Ekman & Heider, 1988; Fridlund, Ekman, & Oster, 1987). Next, four emotion labels that characterize "milder" forms of basic emotions were selected. For example, "irritated" was chosen as a less intense version of "angry" and "amused" was chosen to indicate a lower affective level of "happy." Finally, emotions that could be thought of as more complex or higher order emotions were selected. The 15 emotions were randomly placed across various letters on the computer keyboard. In order to become comfortable with the task, participants were given a practice trial that included three movie clips. They were not told ahead of time that some movies contain conflicting emotional cues, although this was explained to them in the debriefing following participation. Movie run order was counterbalanced across participants. Each participant was exposed to 16 runs of 20 movies each.

Figure 3 Sample still images from emotional stimuli (left to right: angry, fearful, happy).

Look in the box.
Clouds are in the sky.
It’s dark already.
The dog is barking.
The door is open.
I didn’t expect you.
It might happen.
Put it down.
It’s across the street.
Turn off the television.

Figure 4 Emotionally ambiguous sentences. Each sentence was recorded in four affective states: angry, fearful, happy, and neutral.

Basic Emotions

Happy
Angry
Fearful
Sad
Surprise
Disgust
Neutral

Mild Emotions

Amused
Irritated
Anxious
Discouraged

Complex Emotions

Sarcastic
Puzzled
Confident
Relief

Figure 5 Emotions available for selection by participants in a forced-choice format.

Results

Preliminary Analyses

Data were double entered into Microsoft Excel for subsequent use in SPSS. Exclusion criteria were examined to identify participants that would be eliminated from statistical analyses. Two participants scored below normal on the Benton Facial Recognition Test and were therefore eliminated from all statistical analyses. Additionally, 8 participants scored greater than a T-score of 70 and met criteria for the Symptom Count Cutoff Score on the ASRI-4. In order to determine if there were any differences in performance on the assessment measures between these participants and participants who were not elevated on the ASRI-4 independent t-tests were performed. Results indicated that there were no significant differences between these two groups, thus these participants were included in the final sample.

The experimental stimuli included a total of 320 movie clips. Eighty contained audio and video tracks with matching emotions (i.e., Happy Face / Happy Voice) and the remaining 240 stimuli contained incongruent emotional cues (i.e., Happy Face / Angry Voice). Frequency of responses for the 240 incongruent stimuli can be seen in Table 1. The maximum number of times a response could be chosen for each condition is 740: (20 movie clips per condition) x (37 participants). Twenty-seven data points (.91%) were missing for the congruent stimuli. Fifty-two data points (.59%) were missing for the incongruent stimuli, with missing data points in any one condition being no more than 1.08%. Total missing data points consisted of .66% of the total data points. The current data set is large and only a small number of random points of data are missing, thus missing data points were handled through casewise deletion in all statistical analyses.

As can be seen in Figure 6, the most frequent response types across all incongruent conditions reflected the emotions presented in the stimuli (Angry, Fearful, Happy, and Neutral).

The most frequent responses after those were Irritated, Anxious, and Puzzled. See Figures 7-10 for frequency tables for congruent conditions; see Figures 11- 22 for frequency tables for incongruent conditions. Frequency of responses for each incongruent condition can be seen in Table 1.

Primary Analyses

Hypothesis 1: Statistical analyses were performed in order to determine if there was a significant difference between reaction times for congruent and incongruent emotional stimuli. Sample sizes were unequal, thus a Wilcox signed ranks test and a Sign test were performed in addition to a paired t-test. Results did not differ between the nonparametric and parametric analyses; therefore, the paired t-test results are reported. Results indicate that participants took significantly longer to respond to incongruent emotional stimuli (M = 3640.15 ms, SD = 994.33) than to congruent emotional stimuli (M = 3035.75 ms, SD = 839.42), $t(36) = -10.39$, $p = .000$.

Hypothesis 2: A confusion matrix was created to test the hypothesis that when presented with matching emotional stimuli, participants would choose the correct emotion (Table 2). Percentages of correct inference can be seen through the percentages in the diagonal of the matrix as well as the pattern of errors or confusions in the off-diagonal entries. The diagonal components of the confusion matrix reveal that the emotions portrayed in the congruent stimuli can be recognized by participants with more than 82 percent accuracy. The recognition rates for Angry, Happy, Fearful and Neutral utterances were all fairly high (ranging from 82% to 95%). According to the confusion matrix the most easily recognizable category is Happy (95.4%) and the least easily recognizable category is Neutral (82.2%). Additional response choices were not included in the confusion matrix, as this analysis only takes into account correct and incorrect responses; thus, only responses which were represented by the stimuli were included.

In order to take other responses into account, sign tests were also conducted in order to determine whether participants correctly chose the emotion presented in the matching emotional stimuli. Analyses were conducted between participants who chose the correct emotion as their modal response for the 20 movies in the congruent conditions and participants who chose an incorrect emotion as their modal response. Sign tests indicated that for all matching conditions participants chose the correct response significantly more often then they chose an incorrect response (Table 3).

Hypothesis 3: A series of separate sign tests were conducted to determine if during incongruent conditions, participants tended to perceive the emotion presented in the face more often than the emotion presented in the voice. The sum of participant responses that were reflective of the emotion presented in the face were compared to the sum of participant responses reflective of the emotion presented in the voice (Table 4). Sign tests indicated that in all but two conditions (Neutral Face / Angry Voice; Neutral Face / Fearful Voice) participants were more likely to choose a response consistent with the emotion portrayed in the face. Although these two conditions were not significantly different, the pattern of responses reflected a bias toward the emotion presented in the voice. In order to further explore the observed patterns particular emotion responses were collapsed into one response category. Irritated was thought to be an emotional response representative of a milder form of Angry. Thus, a sign test was also performed combining the sum of Angry and Irritated responses for the Neutral Face / Angry Voice condition. Analyses indicated that when these two response choices were combined, participants were more likely to choose a response consistent with the emotion portrayed in the voice. Similarly, Anxious was included as an emotional response representative of a milder form of Fearful. Thus, a sign test was also performed combining the sum of Fearful and Anxious responses for the Neutral Face / Fearful Voice condition. Analyses indicated that when Fearful and Anxious responses were combined there was a tendency for participants to select responses reflective of the emotion presented in the voice, p = .067 (See Table 5).

Sign tests utilizing participant modal responses were also conducted. Results utilizing participant modal responses were similar to those using the sum of participant responses. Analyses utilizing participant modal responses indicated that in all but three conditions (Fearful Face / Angry Voice; Neutral Face / Angry Voice; Neutral Face / Fearful Voice) participants were more likely to choose a response consistent with the emotion portrayed in the face. The only condition that differed from the results found when using the sum of participant responses was the Fearful Face / Angry Voice condition.

As was done with the sum of participant responses, a sign test combining Angry and Irritated responses for the Neutral Face / Angry Voice Condition was conducted. Analyses using participant modal responses also indicated that when these two response choices were combined participants were more likely to choose a response consistent with the emotion portrayed in the voice. Similarly, a sign test was also performed combining Fearful and Anxious responses for the Neutral Face / Fearful Voice condition. Results using participant modal responses indicated that even after combing Fearful and Anxious responses, this condition was not significantly biased towards the emotion presented in the face or voice. In order to further examine the Fearful Face / Angry Voice condition, Fearful and Anxious responses were combined and Angry and Irritated responses were combined. Sign tests for both the sum of participant responses and modal responses were not significant when responses were collapsed across these response choices (See Table 5).

Certain emotions have similar acoustic parameters which may make them difficult to classify or more likely to be mutually confused. (De Silva, Miyasato, & Nakatsu, 1997). For example, in neutral and sad sentences the energy and the pitch are usually maintained at the same level. Therefore, these emotions are difficult to classify when presented aurally. Moreover, sadness is typically considered an emotional category that is negative in valence but low in arousal, similar to the way a neutral presentation is perceived (Russell, 1980). Thus, the presentation of sadness is quite different from the presentation of fear. Based on this information Neutral and Sad responses were collapsed for exploratory analyses. A sign test combining the sum of Neutral and Sad responses for the Neutral Face / Fearful Voice condition indicated that participants were more likely to choose a response consistent with the emotion portrayed in the face when these two response choices were combined. However, when utilizing participant modal responses, when Neutral and Sad were combined for the Neutral Face / Fearful Voice condition, results were not significantly biased towards the emotion presented in the face or voice. A sign test was also performed for the sum of responses and modal responses in which Neutral and Sad were collapsed into one category and Fearful and Anxious were collapsed into a second category. Although not significant for the modal responses, a sign test utilizing the sum of responses indicated that there was a tendency for participants to select responses reflective of the emotion presented in the face (See Table 5). Although it seems that more participants choose a response reflective of the emotion presented in the face when both of these categories are collapsed, it is difficult to make specific conclusions about this condition until more information is garnered regarding the distinction between Neutral and Sad facial expressions in the experimental stimuli used in the current study.

Hypothesis 4: According to this hypothesis, the sample would be divided based on individual preferences or individual differences in emotion perception and therefore would not show an overall group bias towards a single sensory modality. In addition, this hypothesis predicted that there may be modality preference that is dependent upon group membership (i.e. sex). A Kolmogorov-Smirnov Two-Sample Test was utilized to determine if there was a difference in modal response type between male and female participants. Results indicated that there is no significant difference in modal response type between male and female participants, $p = .990$. Despite the lack of group differences between sexes, individual preferences in response choice were examined. Analyses of modal response type across all conditions revealed that the majority of the sample responded with a modal response reflective of the emotion presented in the face ($n = 30$). Two participants had modal responses reflective of the emotion presented in the voice and 5 participants had modal responses reflective of an emotion other than that expressed in the face or voice. Thus, despite some individual difference in modal response type, the modal response across the sample was reflective of a bias toward the emotion presented in the face. (Table 6).

Hypothesis 5: It was predicted that certain incongruent conditions would elicit integration of the auditory and visual information, resulting in the perception of a third, more complex emotion. In order to determine the percentage of participants who chose an emotion other than that presented in the face or voice, participant responses other than Angry, Fearful, Happy and Neutral were summed across conditions. Despite the salience of facial expressions across conditions, particular incongruent conditions elicited a large percentage of response choices reflective of emotions other than those presented in the face or voice (Table 7).

Ancillary analyses collapsed participant responses based on the positive or negative valence of the chosen response. Response choices that were collapsed into a variable reflecting positive emotions included: Surprised, Amused, Confident, and Relief. Response choices that were collapsed into a variable reflecting negative variables included: Irritated, Sarcastic, Anxious, Sad, Puzzled, Discouraged, and Disgusted. There were an uneven number of possible positive and negative response choices for these analyses. This uneven division of positive and negative response choices directly mimics the uneven grouping of positive and negative response choices across the study (e.g. negative response choices = 9; positive response choices = 5; neutral can be interpreted as being a member of either category).

Sign tests were conducted in order to determine if participants were more likely to choose a response based on the valence of the facial expression in incongruent conditions. Results of the various sign tests are consistent with the idea that when participants chose a response other than that expressed in the face, they were more likely to choose an emotion that was similar to the general affective tone portrayed by the face. However, there are several exceptions to this pattern of results. In the Neutral Face / Angry Voice condition and in the Neutral Face / Fearful Voice condition, sign tests indicated that participants tended to chose an emotion response consistent with the emotion presented by the voice. This could be due to the salience of both anger and fear when compared to a neutral facial presentation. Therefore, in these two conditions, the auditory presentation was more influential in determining participant responses than the presentation in the face.

On the other hand, although a neutral presentation is often thought of as being negative, a neutral presentation can be categorized as either positive or negative. Thus, the higher rate of negative responses in these two conditions could have been determined by either the face (if a neutral presentation was considered to be negative in valence) or the voice. In fact, in the Neutral Face / Happy Voice condition sign tests indicated that participants were more likely to chose a response with a negative valence. Thus, in this particular condition it seems as if a neutral facial presentation is being interpreted as inherently negative. When a Happy Face was paired with a Neutral Voice however, participants were more likely to choose emotion responses with a positive valence.

Although conditions with a Happy Face / Angry Voice and Happy Face / Fearful voice were not significantly different in terms of the positive or negative valence of emotion responses, the pattern of responses indicated that responses with a positive valence were chosen more often than responses with a negative valence lending support to the idea that emotion choices are driven by the valence established by the facial expression. Moreover, incongruent conditions with a negative facial expression resulted in a higher frequency of negative responses.

It is important to note that for conditions that included both angry and fearful in either the face or voice we are unable to determine which modality is driving the response pattern since both of these emotions are perceived as being negative. The overall pattern of responses reflects the idea that when participants did not choose the emotion portrayed by the face, they were still likely to choose a response congruent with the general affective tone portrayed through the facial expression (See Table 8).

Exploratory Analyses

To further explore responses reflective of a response other than that presented in the face or voice, a series of separate qualitative analyses was performed. Graphs were constructed to explore specific combinations of emotions that tended to produce the highest frequency of responses reflective of emotions other than those presented in the face or voice. Responses that resulted in a pattern across conditions or had a high response frequency will be discussed.

Puzzled

Incongruent conditions with a Fearful Face tended to produce the highest frequency of Puzzled responses (Figure 23).

Sarcastic

The specific combination of a Happy Face with Angry Voice elicited the highest frequency of Sarcastic responses (Figure 24).

Surprised

The specific combination of a Fearful Face with Happy Voice elicited the highest frequency of Surprised responses (Figure 25). The conditions displaying a Fearful Face with Neutral Voice and a Fearful Face with an Angry Voice also were likely to elicit Surprised responses. There may be something inherent in a fearful face that signifies and is primarily accounting for the Surprised responses.

Sad

The specific combination of a Neutral Face with Fearful Voice elicited the highest frequency of Sad responses (Figure 26). A high number of Sad responses were also seen in the Angry Face with Fearful Voice condition. These results may suggest that there is something inherent in the acoustic parameters of vocal fear that are similar to or can imply sadness when combined with an incongruent facial expression with a negative or neutral valence.

Confident

When participants were presented with the incongruent combination of an Angry Face with Neutral Voice or a Neutral Face with Angry Voice they tended to interpret the emotion as Confident (Figure 27). Despite the high frequency of responses in both conditions, the response choice Confident was chosen most often in the condition with an Angry Face with a Neutral

Voice. In addition, the conditions representing a Happy Face with an Angry Voice and a Happy Face with a Neutral voice elicited a high number of Confident responses. The specific combinations that elicited Confident responses could reflect the participant's interpretation of the word "Confident." Confident tends to imply two meanings, one positive and the other neutral or negative. For example, confident can be thought of as a feeling one has when they are certain they will do well. On the other hand, confidence can also imply unwarranted faith in oneself or ones abilities. It can be seen as intimidating for someone who lacks confidence. Therefore, participant interpretations of the word itself could account for the conditions with high response rates. The conditions including Angry and Neutral combinations could suggest the more negative connotation of confidence whereas the combinations with a Happy Face could imply the more positive interpretation of confidence.

Amused

Amused was included in the response choices to represent a milder form of Happy. This response appeared more frequently in conditions with a Happy Face, rather than in conditions with a Happy Voice (Figure 28). In fact, it is interesting to note that an Amused response almost never occurred with any other conditions. It has been argued that Happy is a difficult emotion to recognize when depicted in prosody which is in opposition to the often high level of visual recognition.

.Table 1 Frequency of Response for Incongruent Conditions

	Response														
Stimulus (Face – Voice)	Angry	Fearful	Happy	Neutral	Irritated	Surprised	Amused	Sarcastic	Confident	Puzzled	Anxious	Sad	Discouraged	Disgusted	Relief
Angry – Fearful	**279**	132	4	17	87	4	2	6	16	13	50	76	18	31	1
Angry – Happy	**284**	21	14	47	185	16	3	25	27	21	22	4	12	53	2
Angry – Neutral	**348**	21	3	54	133	1	0	9	57	13	8	13	30	48	2
Fearful – Angry	89	**185**	8	17	114	54	2	29	8	83	42	12	21	71	1
Fearful – Happy	19	**197**	12	33	60	112	6	16	6	124	65	27	23	31	4
Fearful – Neutral	23	**241**	5	65	33	65	2	18	10	111	62	22	42	35	2
Happy – Angry	27	4	**378**	30	38	24	**77**	70	40	3	14	0	3	8	18
Happy – Fearful	4	39	**350**	9	9	44	110	18	14	3	49	25	4	4	50
Happy – Neutral	10	5	**434**	59	15	23	63	27	40	3	12	5	4	3	32
Neutral – Angry	183	6	5	**199**	178	5	0	42	50	3	3	5	15	40	1
Neutral – Fearful	14	**203**	4	172	23	10	3	8	5	10	65	175	34	7	4
Neutral – Happy	18	17	41	**355**	87	41	6	44	29	24	30	15	13	12	4

Note. Numbers in bold represent the most frequent responses for each stimulus condition.

Table 2 Confusion Matrix Showing the Percent of Judges Responding with the Labels "Angry," "Fearful," "Happy," and "Neutral" to the Stimulus Intended to Portray Each Emotion

	Response			
Stimulus	Angry	Fearful	Happy	Neutral
Angry	**83.30**	0.00	0.00	0.00
Fearful	0.00	**91.00**	0.00	0.00
Happy	0.00	0.00	**95.40**	0.00
Neutral	0.01	0.00	0.00	**82.20**

Note. The values in the diagonal cells of the confusion matrix, representing recognition accuracy, are in bold.

Table 3 Comparison of Responses for Congruent Emotional Stimuli Based on the Modal Response for Each Condition

Stimulus	Participant Responses			Direction of Difference	*p*-Value
	Correct	*Incorrect*	*Tie*		
Angry Face / Angry Voice	35	2	0	$X_C > X_I$	.000
Fearful Face / Fearful Voice	33	4	0	$X_C > X_I$	.000
Happy Face / Happy Voice	31	6	0	$X_C > X_I$	.000
Neutral Face / Neutral Voice	32	5	0	$X_C > X_I$	.000

Note. C = Correct; I = Incorrect

Table 4 Comparison of Auditory and Visual Responses for Incongruent Emotional Stimuli Based on the Sum of Total Responses for Each Condition

Stimulus	Number of Participants with Specified Bias			Direction of Difference	*p*-Value
	Face	*Voice*	*Tie*		
Angry Face / Fearful Voice	27	10	0	$X_F > X_V$	.009
Angry Face / Happy Voice	35	1	1	$X_F > X_V$	.000
Angry Face / Neutral Voice	32	3	2	$X_F > X_V$	.000
Fearful Face / Angry Voice	26	9	2	$X_F > X_V$	.007
Fearful Face / Happy Voice	36	3	0	$X_F > X_V$	.000
Fearful Face / Neutral Voice	27	8	2	$X_F > X_V$	.002
Happy Face / Angry Voice	34	2	1	$X_F > X_V$	.000
Happy Face / Fearful Voice	33	3	1	$X_F > X_V$	.000
Happy Face / Neutral Voice	34	2	1	$X_F > X_V$	.000
Neutral Face / Angry Voice	16	19	2	$X_F < X_V$	.735
Neutral Face / Fearful Voice	15	22	0	$X_F < X_V$	.324
Neutral Face / Happy Voice	32	4	1	$X_F > X_V$	.000

Note. F = Face; V = Voice

Table 5 Comparison of Combined Responses for Incongruent Emotional Stimuli Based on the Sum of Total Responses for Each Condition

Stimulus	Number of Participants with Specified Bias			Direction of Difference	*p*-Value
	Face	*Voice*	*Tie*		
Sum of Responses					
Neutral Face / Angry Voice [a]	10	26	1	$X_F < X_V$	.012
Neutral Face / Fearful Voice [b]	12	24	1	$X_F < X_V$	.067
Neutral Face / Fearful Voice [c]	26	8	3	$X_F > X_V$	.004
Neutral Face / Fearful Voice [d]	23	11	3	ns	.056
Fearful Face / Angry Voice [e]	2	6	29	ns	ns
Modal Responses					
Neutral Face / Angry Voice [a]	9	25	3	$X_F < X_V$	.010
Neutral Face / Fearful Voice [b]	9	13	15	ns	ns
Neutral Face / Fearful Voice [c]	20	12	5	ns	ns
Neutral Face / Fearful Voice [d]	20	13	4	ns	ns
Fearful Face / Angry Voice [e]	17	16	4	ns	ns

Note. [a] Voice response includes both Angry and Irritated responses; [b] Voice response includes both Fearful and Anxious responses; [c] Face response includes both Neutral and Sad responses; [d] Face response includes both Fearful and Anxious responses; Voice response includes both Angry and Irritated responses; [e] Face response included both Neutral and Sad responses and Voice response includes both Fearful and Anxious responses.

Table 6

Individual Modal Responses for Incongruent Emotional Stimuli

Participants	Emotional Stimuli											
	Angry Fearful	Angry Happy	Angry Neutral	Fearful Angry	Fearful Happy	Fearful Neutral	Happy Angry	Happy Fearful	Happy Neutral	Neutral Angry	Neutral Fearful	Neutral Happy
1	2	3	3	2	1	2	2	2	2	2	2	2
2	1	1	1	1	1	1	1	1	1	2	2	2
3	1	3	3	3	3	1	1	3	1	2	3	1
3	3	1	1	2	3	2	1	1	1	2	3	2
3	2	3	2	2	3	2	3	3	2	3	3	1
6	3	3	3	3	3	3	3	1	1	3	2	1
7	1	1	1	3	1	1	1	1	1	1	1	1
8	1	1	1	3	3	1	1	1	1	2	3	1
9	1	1	1	3	3	3	3	3	1	3	3	1
13	1	1	1	1	1	1	1	1	1	2	2	1
11	3	3	3	3	1	3	1	1	1	3	1	1
12	2	3	3	1	1	1	1	1	1	3	3	1
13	1	1	3	1	1	1	1	1	1	1	1	1
13	1	1	1	3	1	1	3	3	1	2	3	3
13	1	1	1	3	3	1	1	1	1	1	1	1
16	2	1	1	3	3	3	3	1	1	3	3	3
17	1	1	1	1	1	1	1	1	1	2	2	1
18	1	1	1	1	1	1	1	1	1	1	1	1
19	3	3	3	3	1	3	1	1	1	3	2	1
23	1	1	1	2	3	3	1	3	1	1	3	1
21	3	3	3	3	3	3	3	3	1	3	2	1
22	1	1	1	1	1	1	1	1	1	1	1	1
23	1	1	1	1	1	1	1	1	1	1	1	1
23	3	3	1	3	3	3	3	3	3	3	3	1
23	1	1	1	2	1	1	2	1	1	2	3	2
26	1	3	3	1	3	3	3	3	1	3	1	3
27	1	1	3	1	1	1	1	1	1	3	2	1
28	1	3	1	3	3	3	1	3	1	1	3	1
29	3	3	1	1	1	1	1	3	1	2	2	3
30	1	1	1	1	1	1	1	1	1	2	2	1
31	1	1	1	3	3	1	1	1	1	2	2	1
32	2	3	3	3	3	1	1	1	1	3	2	1
33	2	3	1	1	1	1	1	1	1	2	2	1
34	1	1	1	3	3	3	1	3	1	2	3	1
35	1	1	1	3	3	3	1	1	1	1	1	1
36	3	3	1	1	3	3	3	1	3	3	3	1
37	3	3	1	3	1	1	1	3	1	3	3	1

Note. 1 = modal response reflective of the emotion presented in the face; 2 = modal response reflective of the emotion presented in the voice; 3 = modal response reflective of an emotion other than the emotion presented in the face or voice.

Table 7 Percentage of Response Choices Reflective of Emotions Other Than Those Presented in the Face or the Voice

	Facial Expression			
Tone of Voice	Angry	Fearful	Happy	Neutral
Angry	_	59.05	39.86	46.22
Fearful	41.08	_	44.59	46.49
Happy	50.00	64.05	_	41.22
Neutral	42.26	54.32	30.68	_

Table 8 Comparison of Positive and Negative Affective Responses for Incongruent Emotional Stimuli Based on the Sum of Responses for Each Condition

Stimulus	Responses			Direction of Difference	*p*-Value
	Positive	*Negative*	*Tie*		
Angry Face / Fearful Voice	1	35	1	$X_P < X_N$	.000
Angry Face / Happy Voice	2	35	0	$X_P < X_N$	.000
Angry Face / Neutral Voice	3	33	1	$X_P < X_N$	.000
Fearful Face / Angry Voice	0	35	2	$X_P < X_N$	.000
Fearful Face / Happy Voice	3	34	2	$X_P < X_N$	.000
Fearful Face / Neutral Voice	2	33	2	$X_P < X_N$	.000
Happy Face / Angry Voice	20	15	2	$X_P > X_N$	.499
Happy Face / Fearful Voice	21	11	5	$X_P > X_N$	.112
Happy Face / Neutral Voice	26	6	5	$X_P > X_N$	.001
Neutral Face / Angry Voice	3	32	2	$X_P < X_N$	.000
Neutral Face / Fearful Voice	1	34	2	$X_P < X_N$	.000
Neutral Face / Happy Voice	7	28	2	$X_P < X_N$	.001

Note. P = Positive; N = Negative

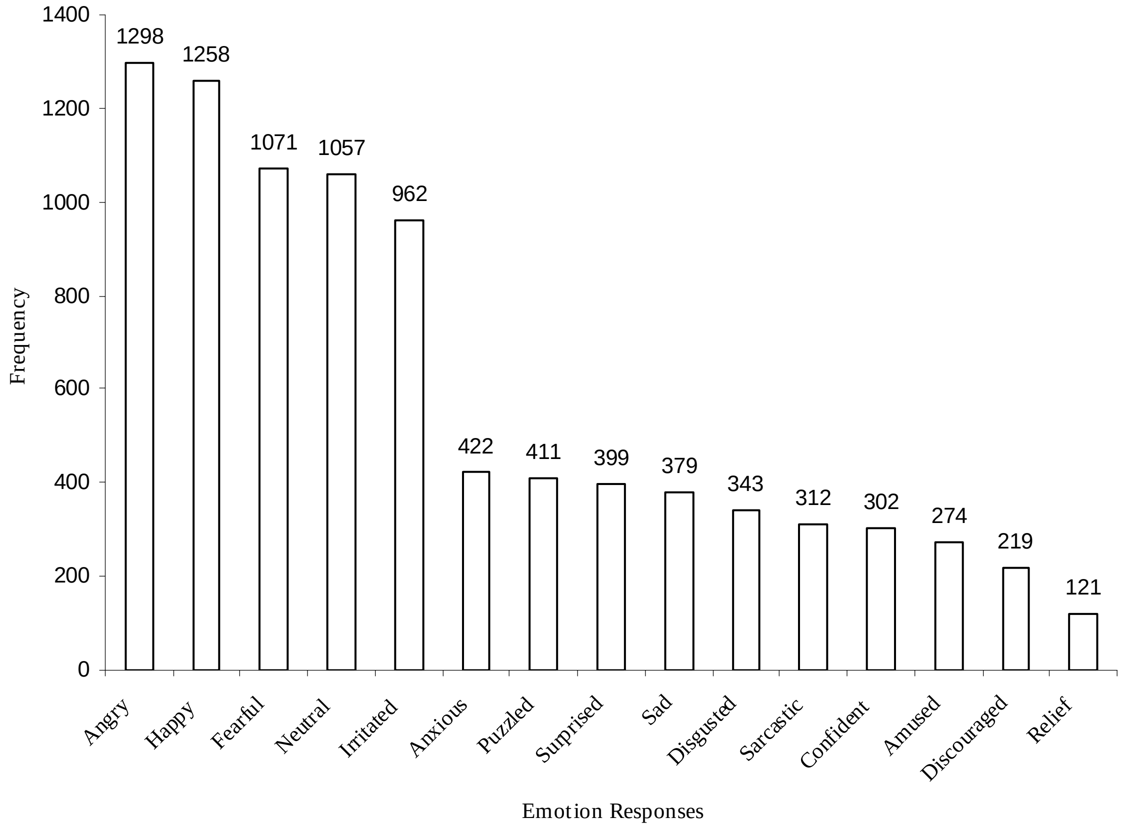

Figure 6 Frequency of Responses across Incongruent Stimulus Conditions.

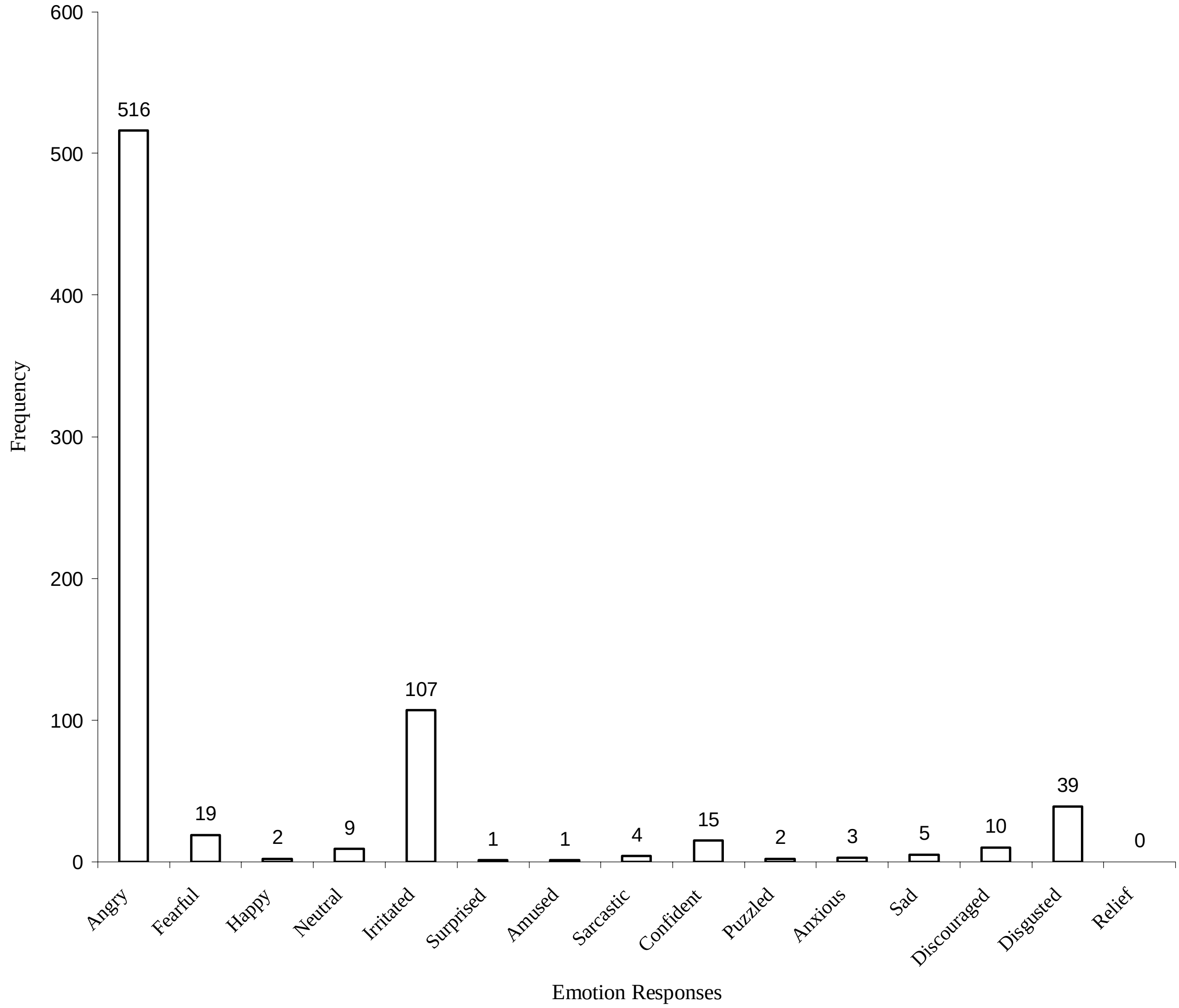

Figure 7 Frequency of Responses for Angry Face / Angry Voice Condition.

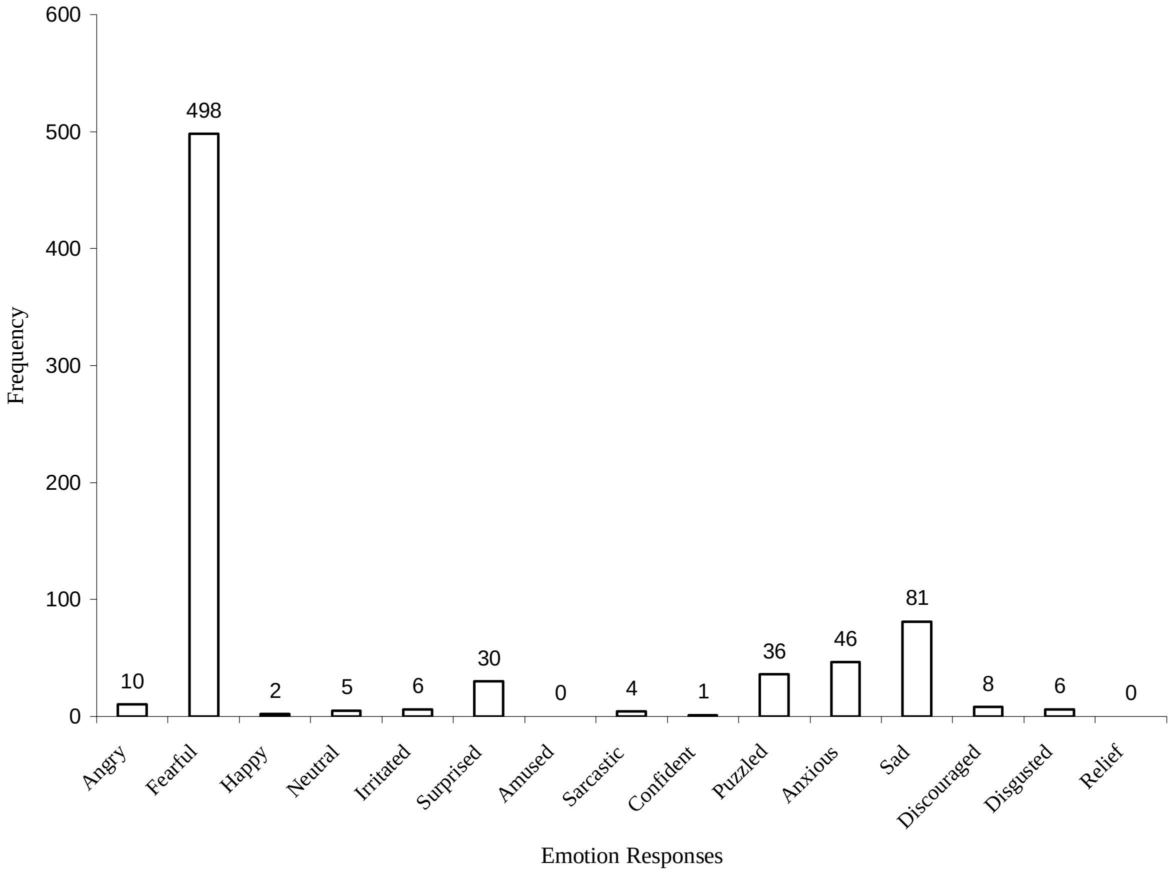

Figure 8 Frequency of Responses for Fearful Face / Fearful Voice Condition.

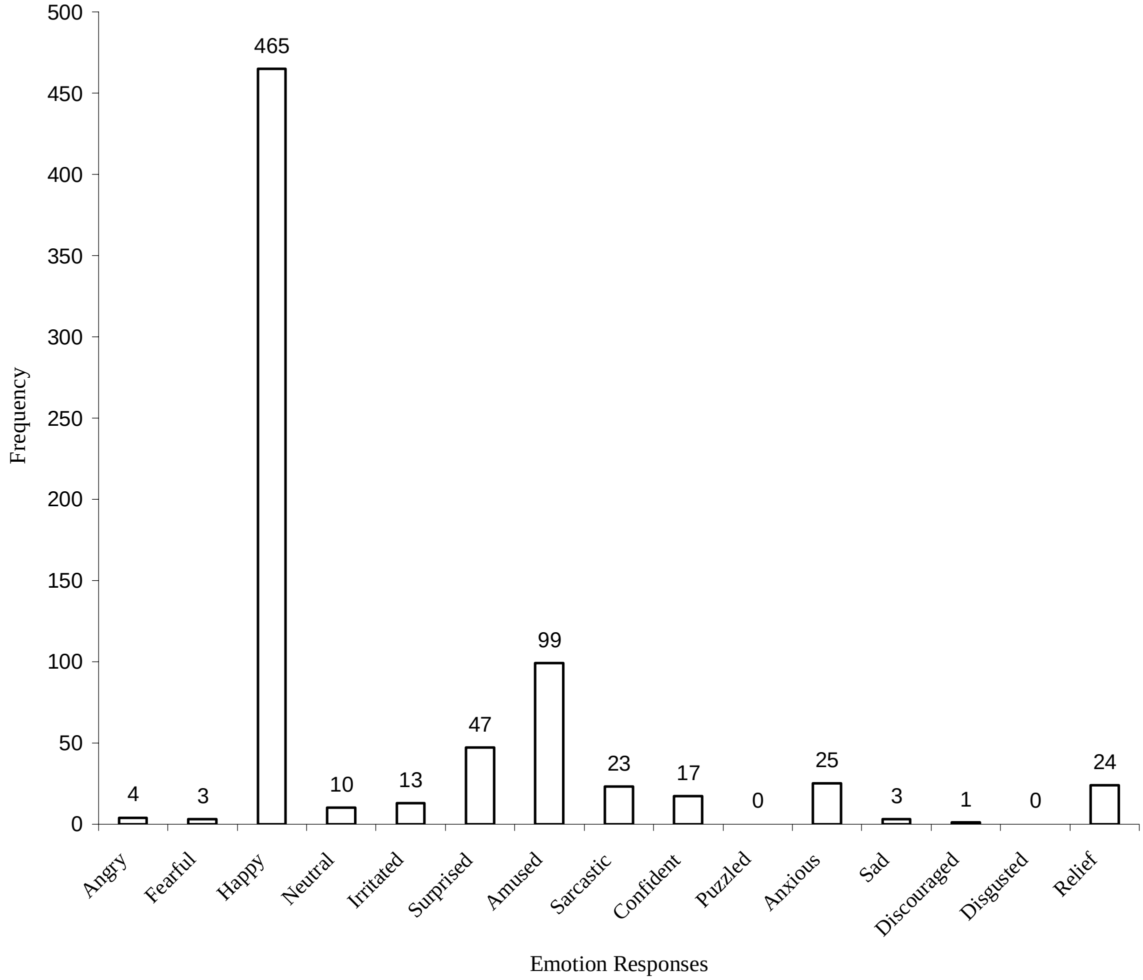

Figure 9 Frequency of Responses for Happy Face / Happy Voice Condition.

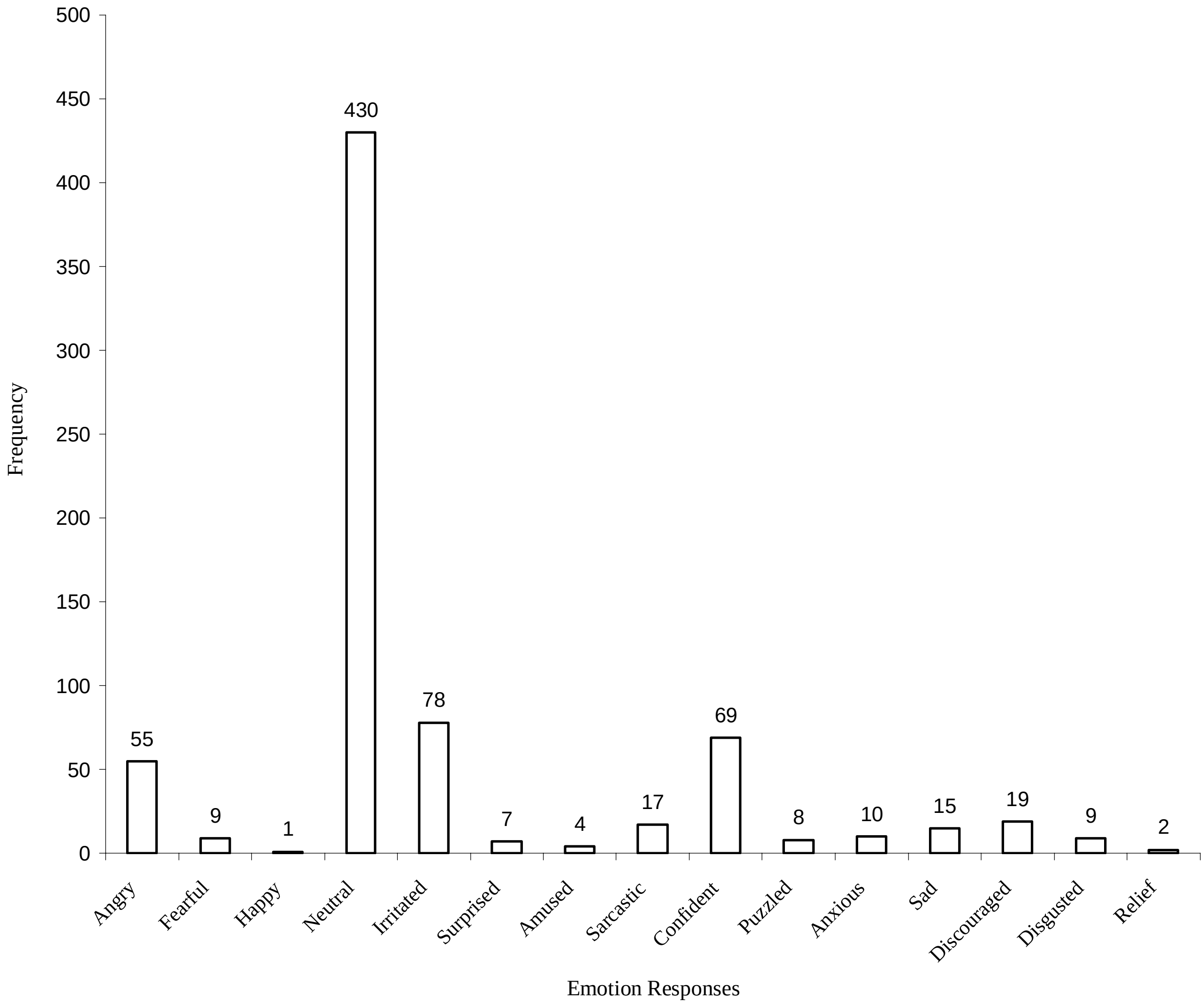

Figure 10 Frequency of Responses for Neutral Face / Neutral Voice Condition.

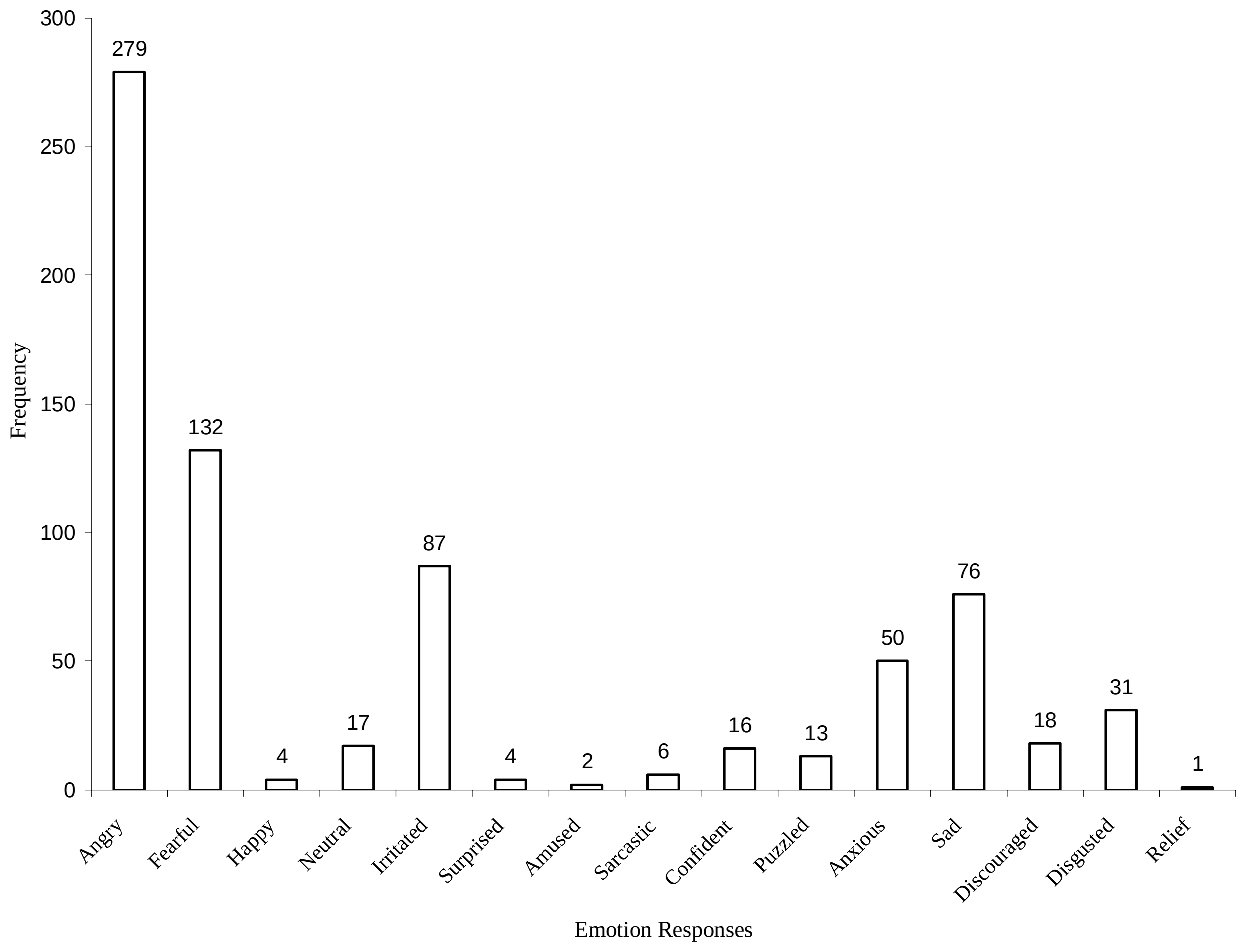

Figure 11 Frequency of Responses for Angry Face / Fearful Voice Condition.

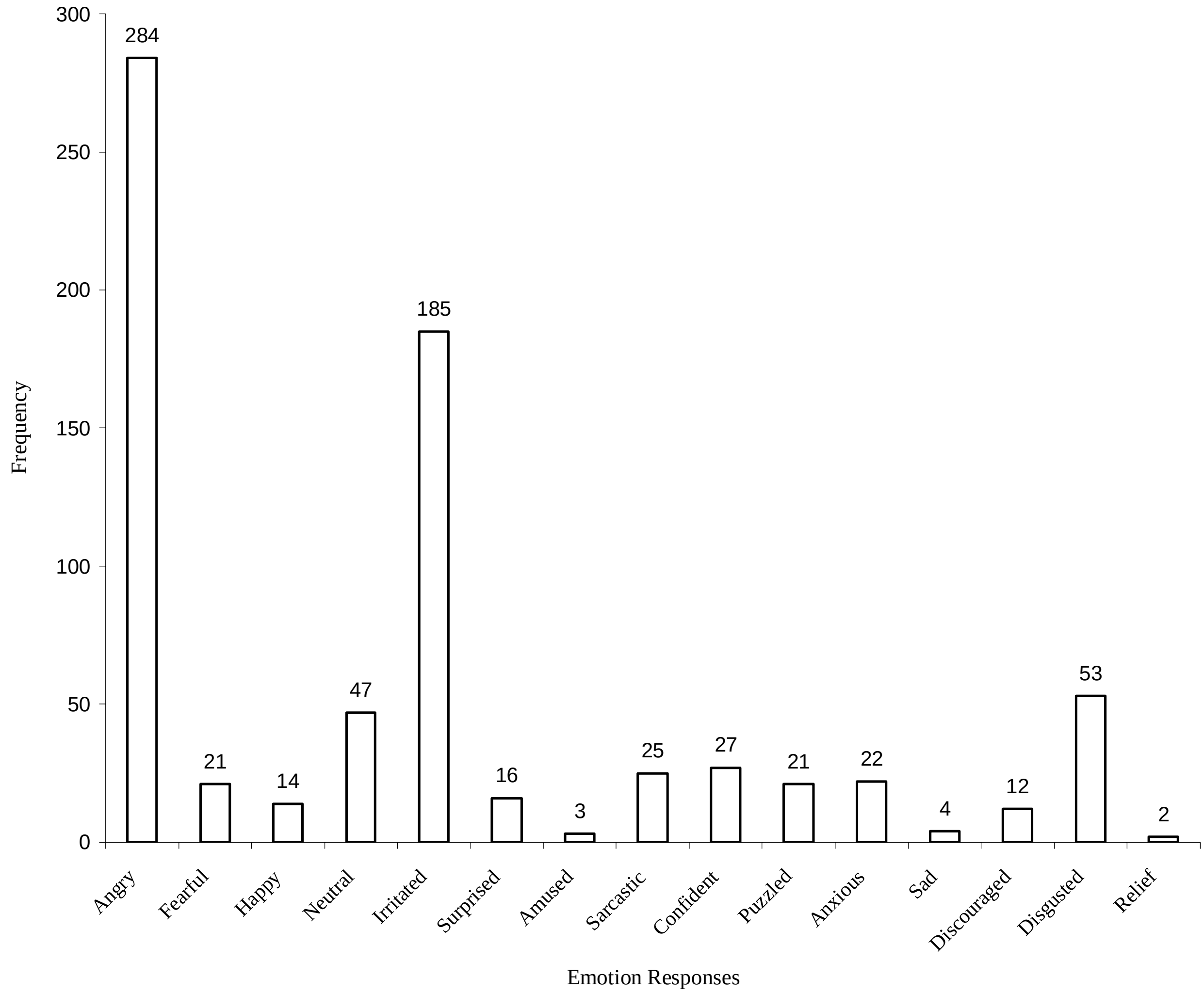

Figure 12 Frequency of Responses for Angry Face / Happy Voice Condition.

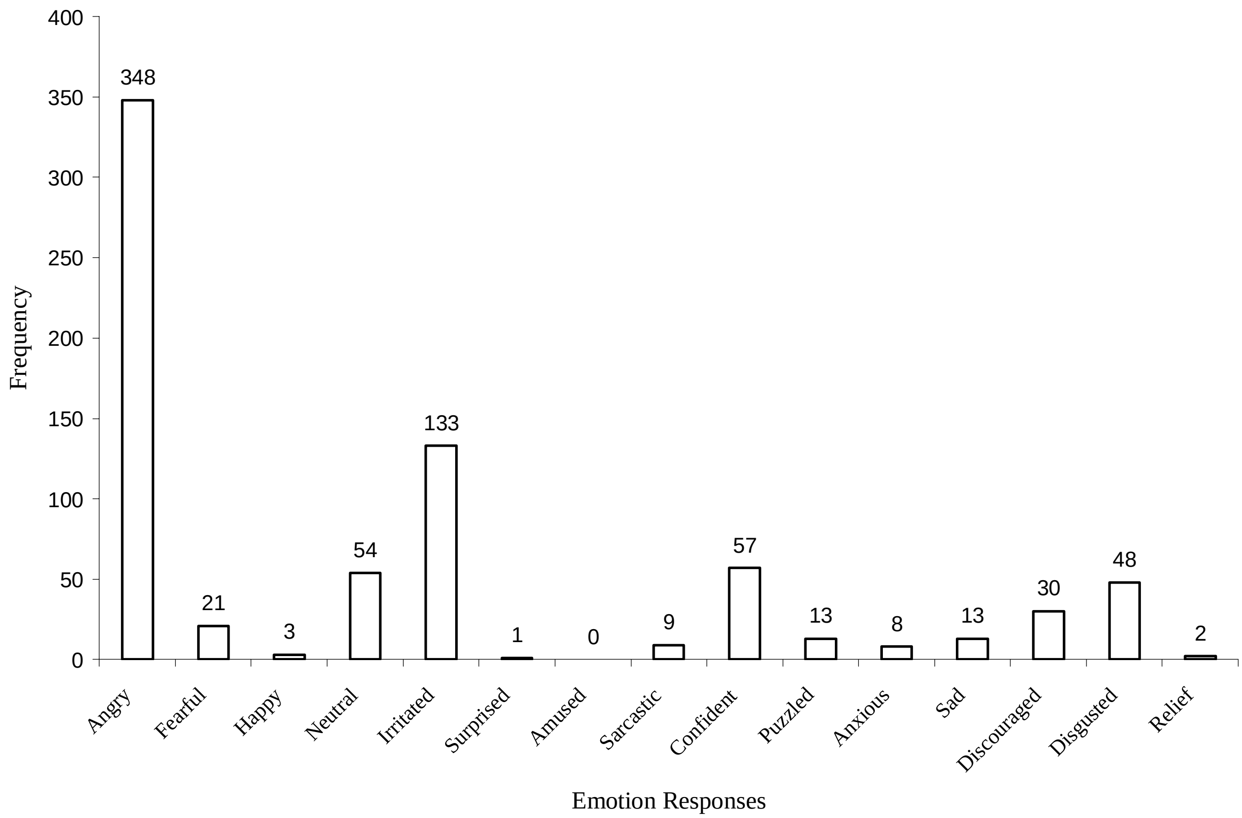

Figure 13 Frequency of Responses for Angry Face / Neutral Voice Condition.

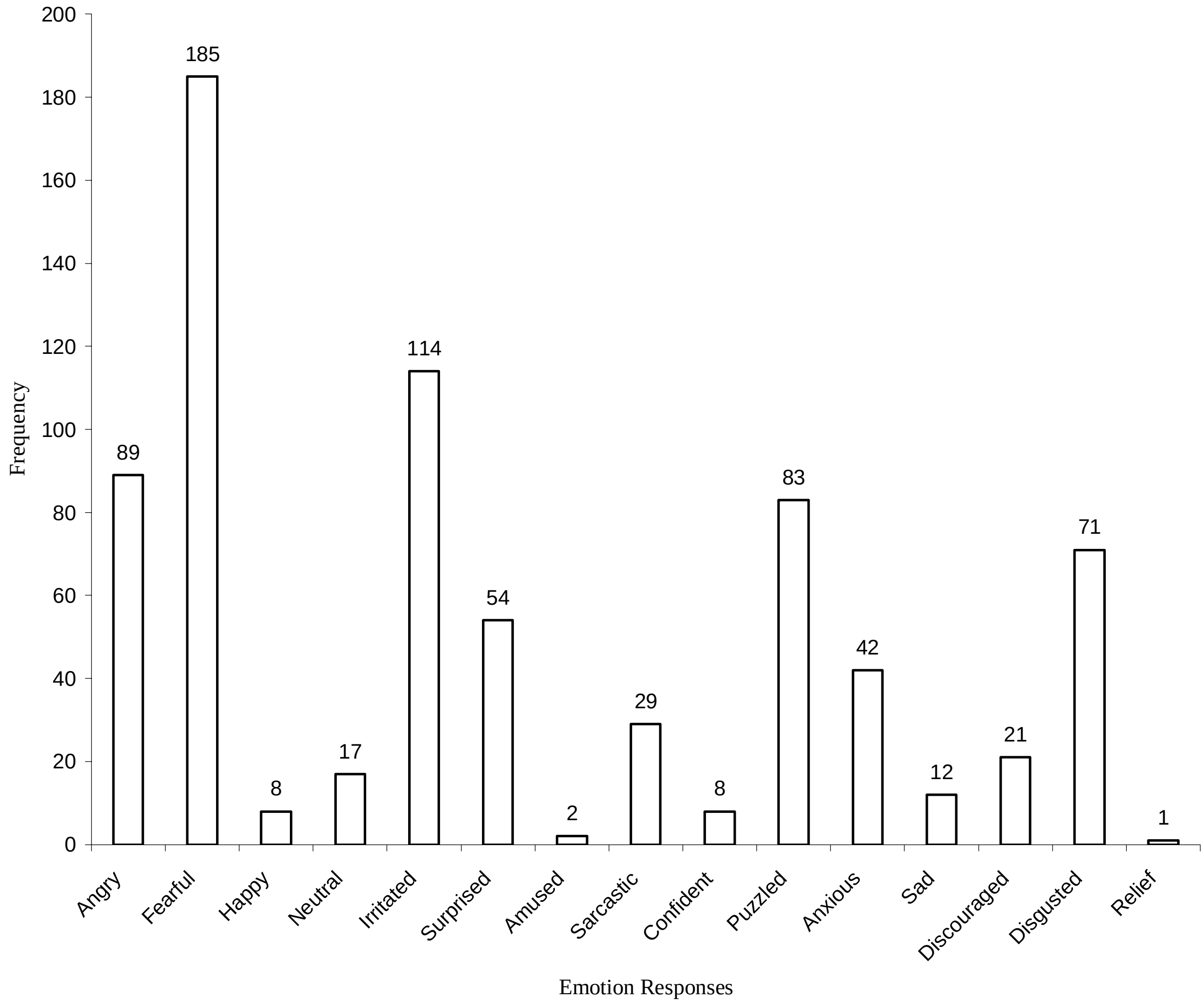

Figure 14 Frequency of Responses for Fearful Face / Angry Voice Condition.

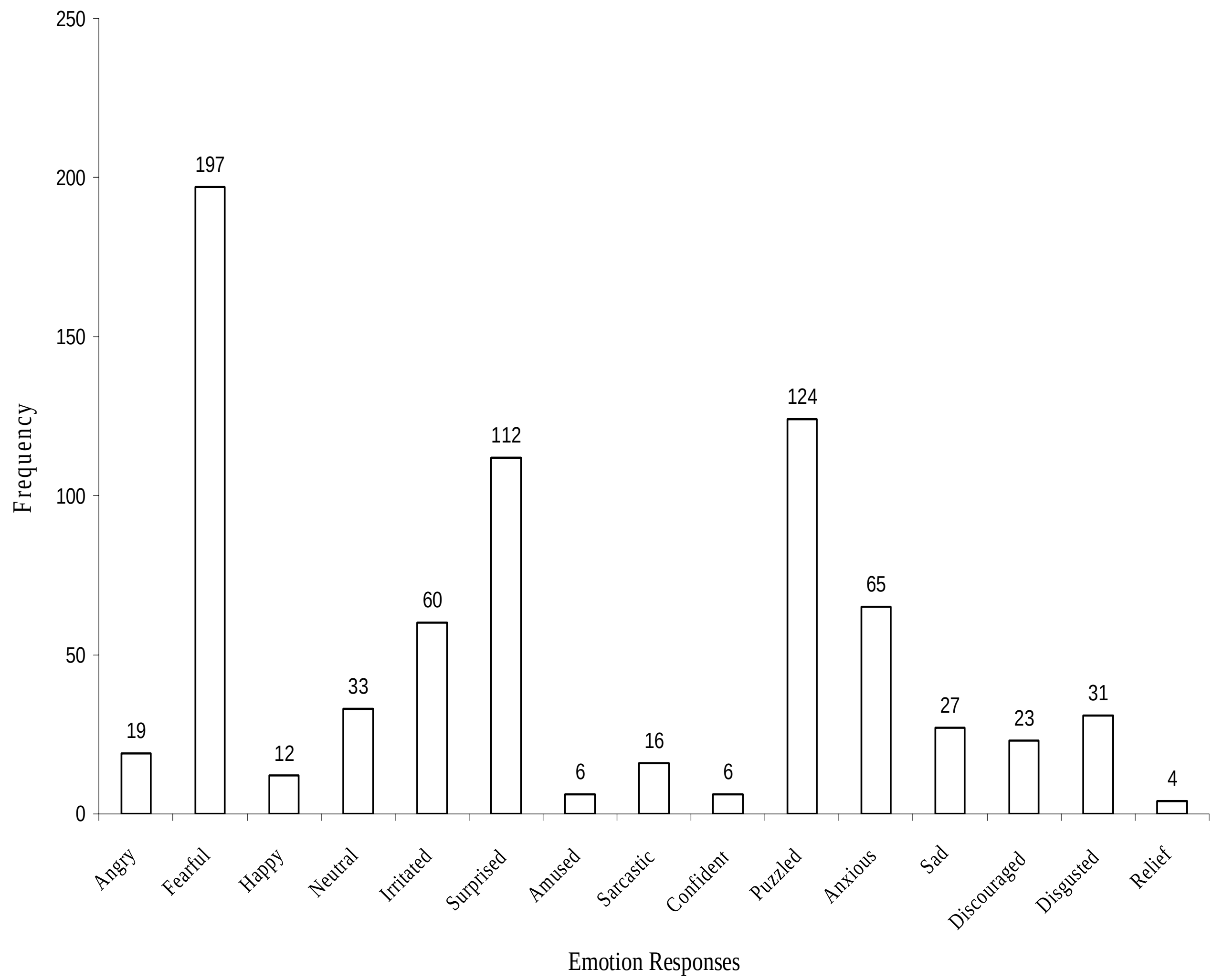

Figure 15 Frequency of Responses for Fearful Face / Happy Voice Condition

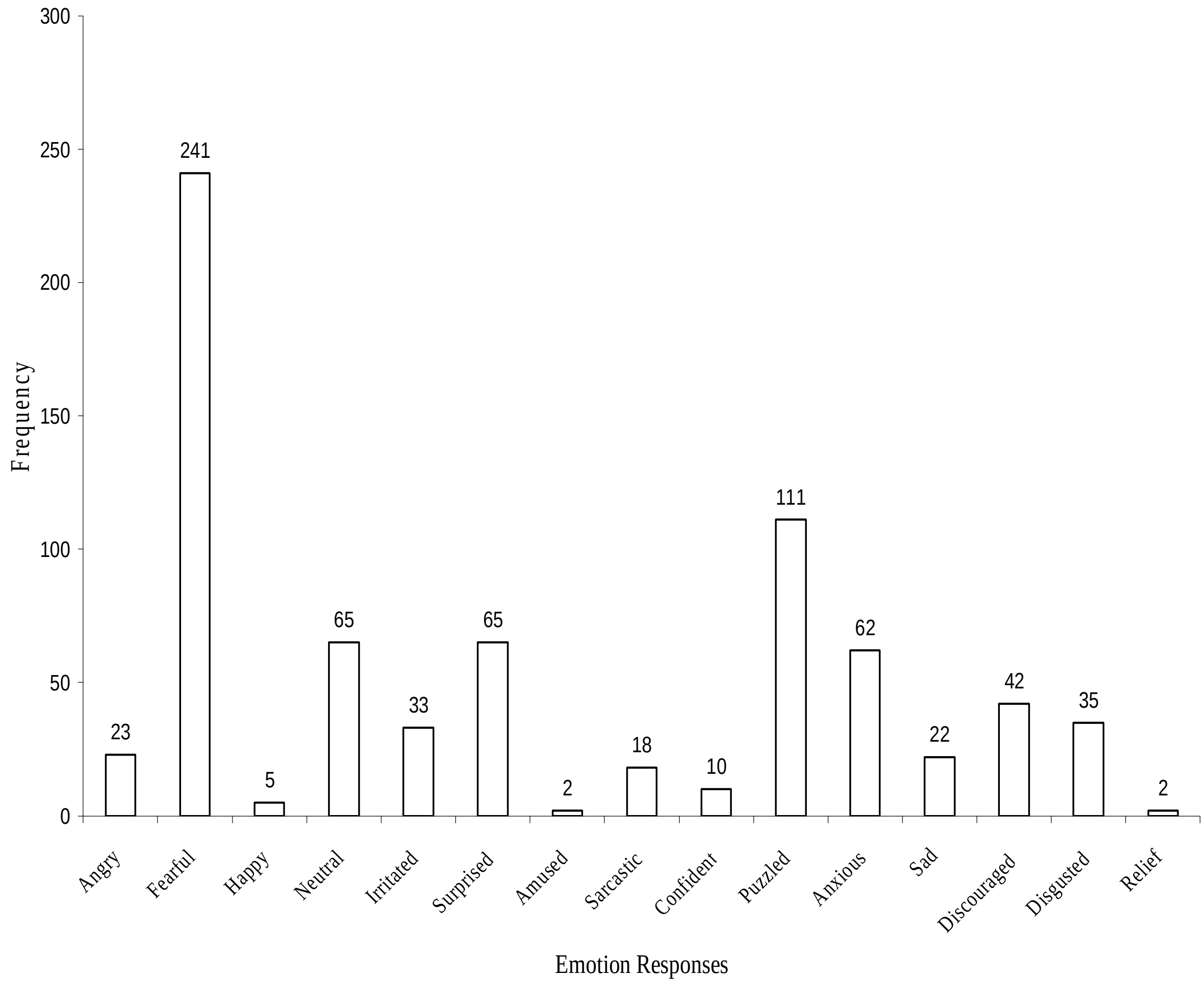

Figure 16 Frequency of Responses for Fearful Face / Neutral Voice Condition.

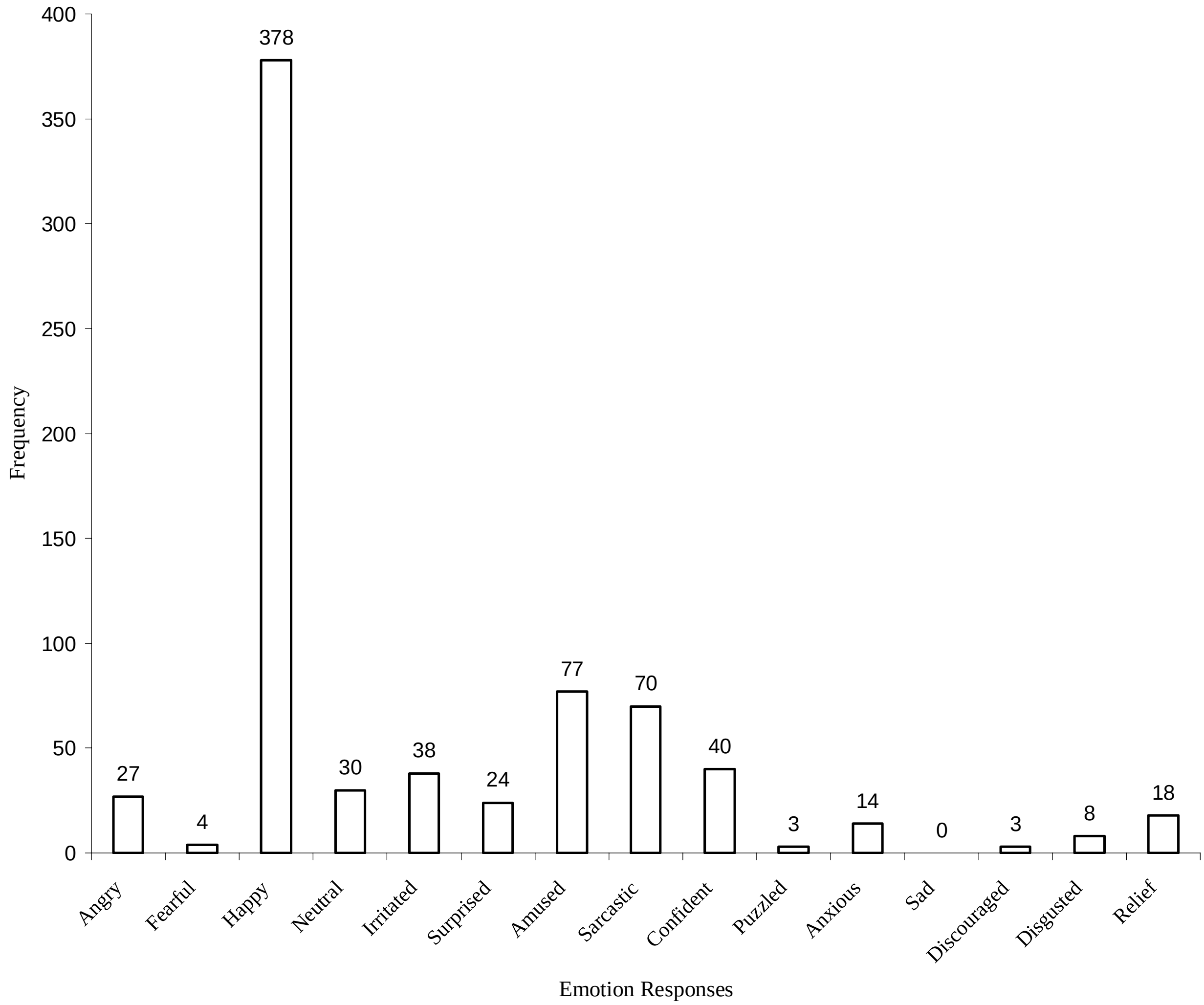

Figure 17 Frequency of Responses for Happy Face / Angry Voice Condition.

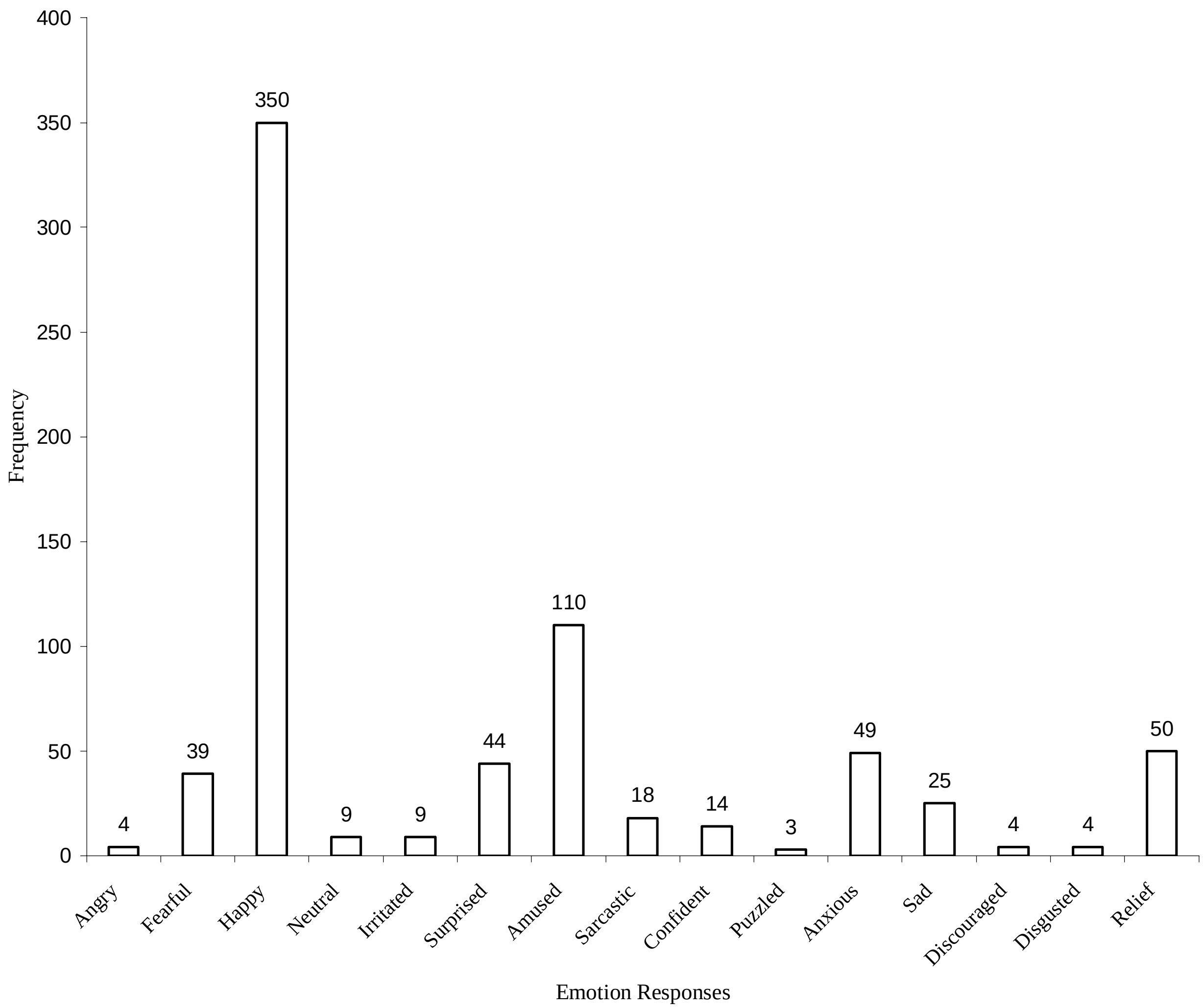

Figure 18 Frequency of Responses for Happy Face / Fearful Voice Condition.

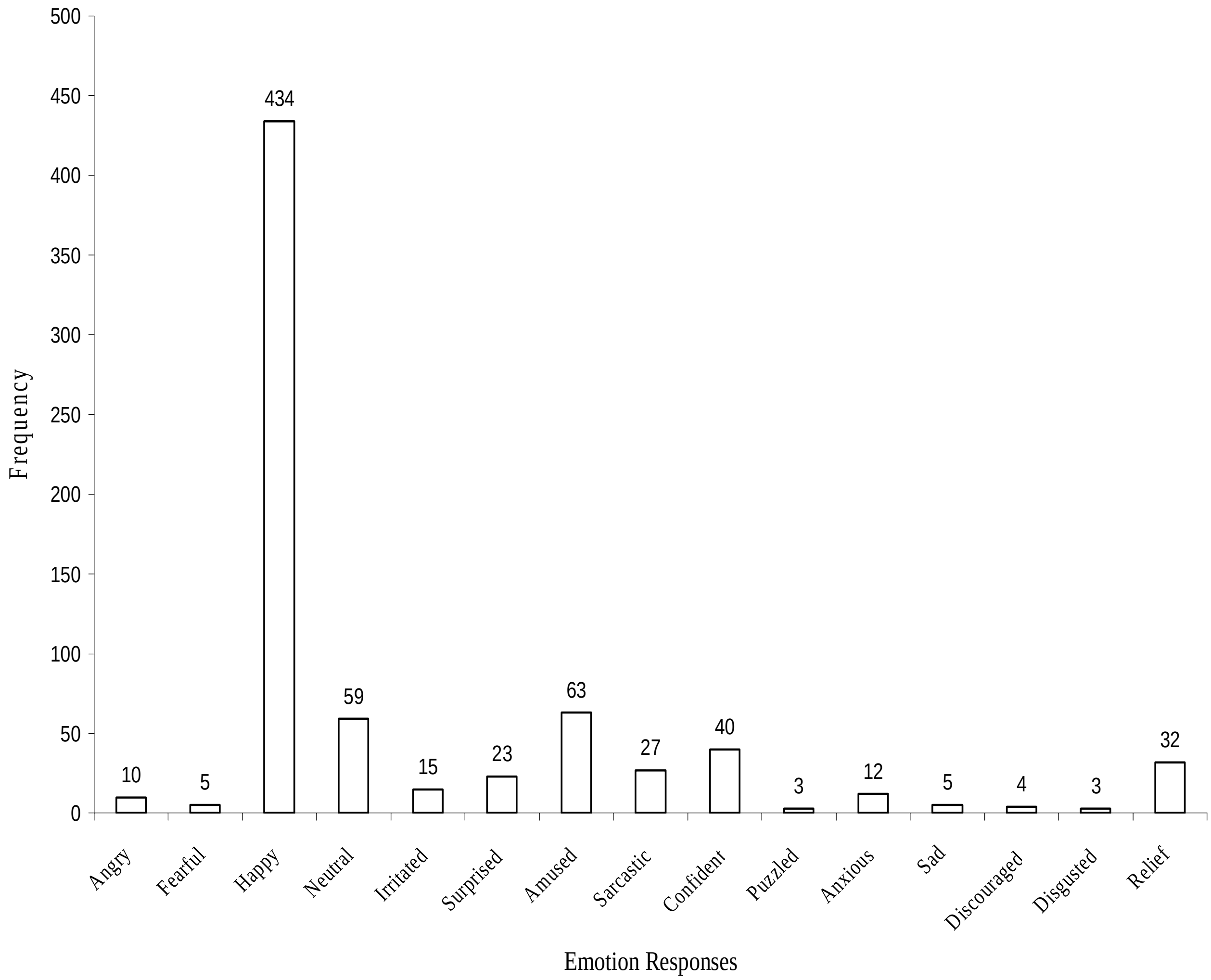

Figure 19 Frequency of Responses for Happy Face / Neutral Voice Condition.

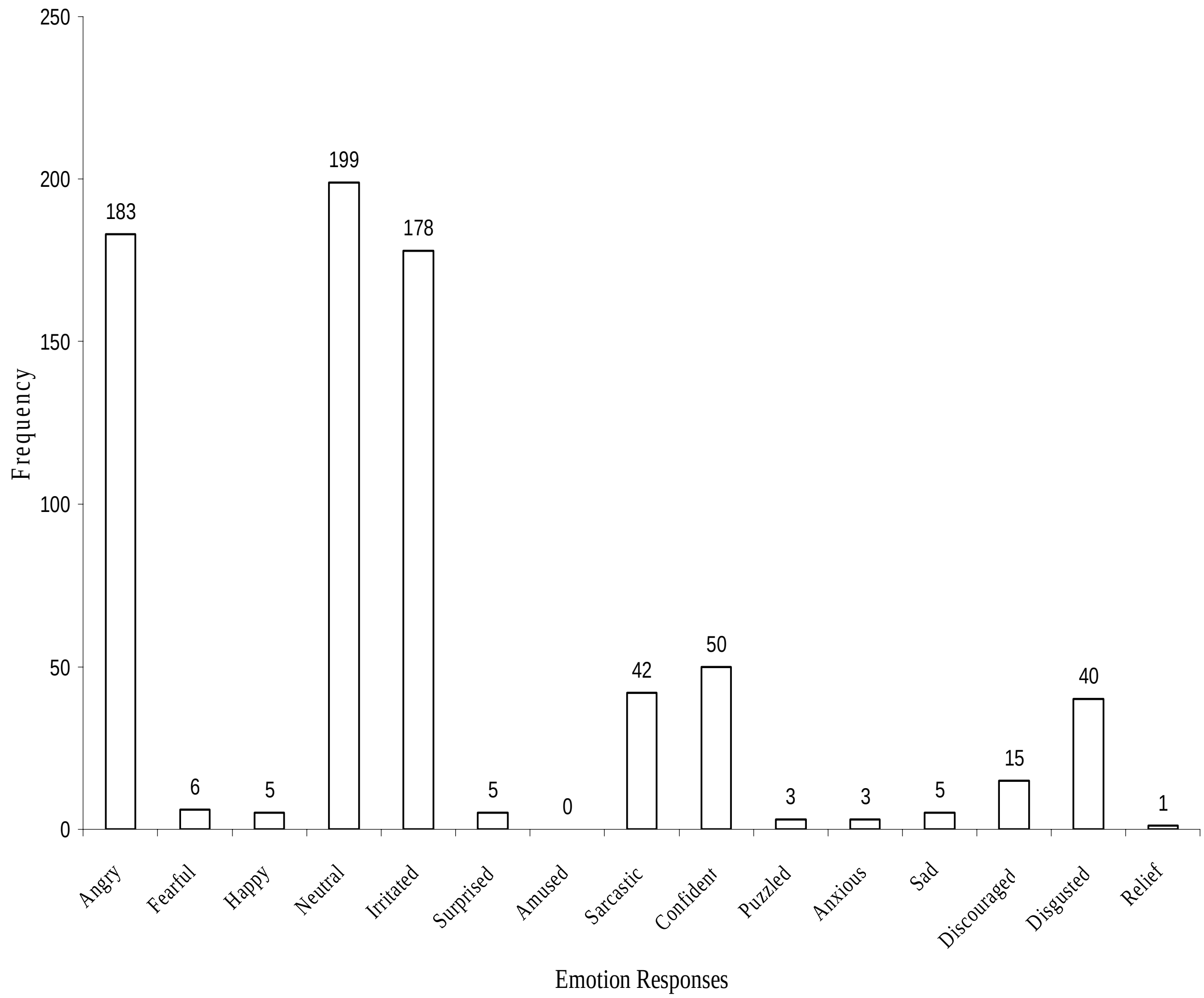

Figure 20 Frequency of Responses for Neutral Face / Angry Voice Condition.

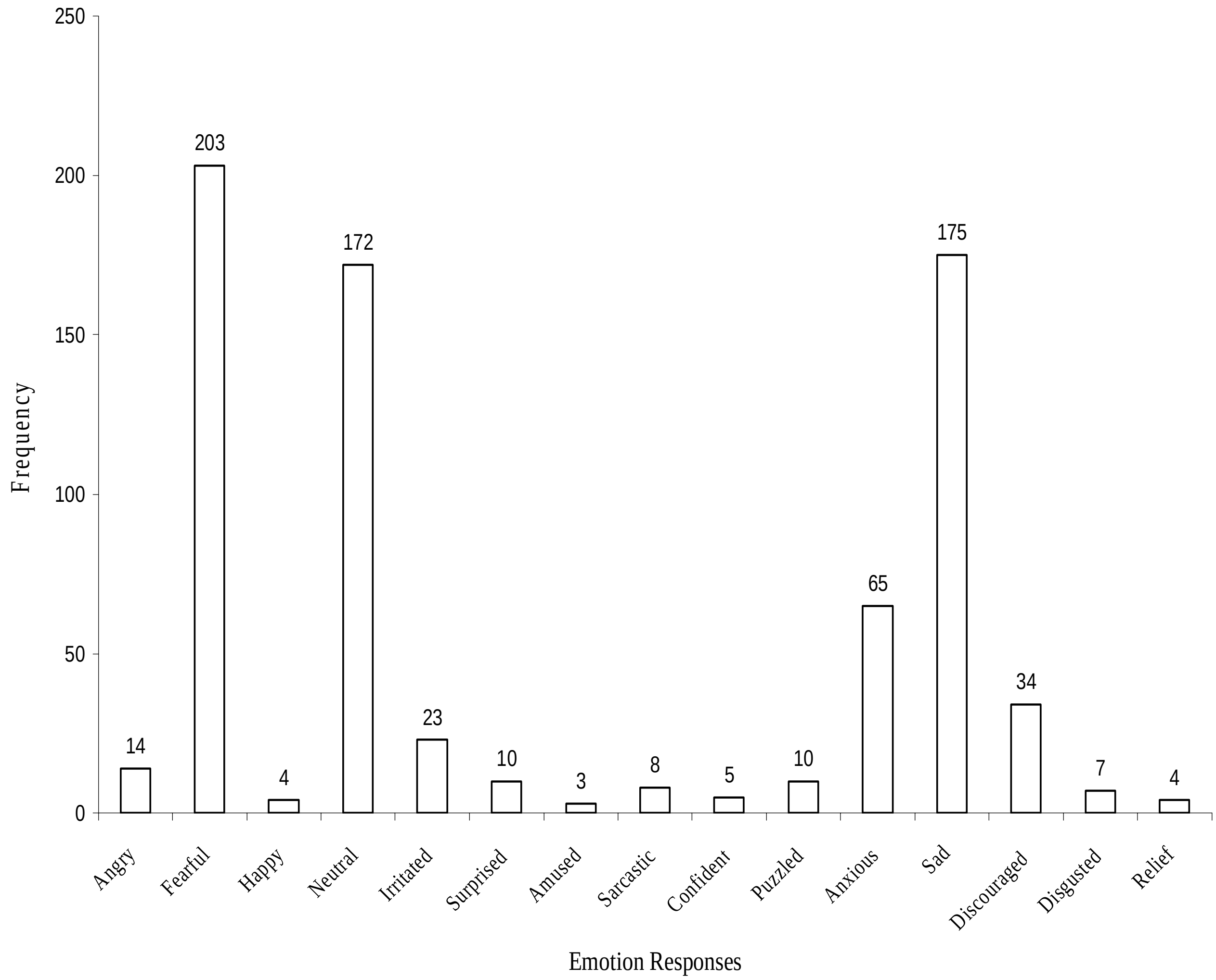

Figure 21 Frequency of Responses for Neutral Face / Fearful Voice Condition.

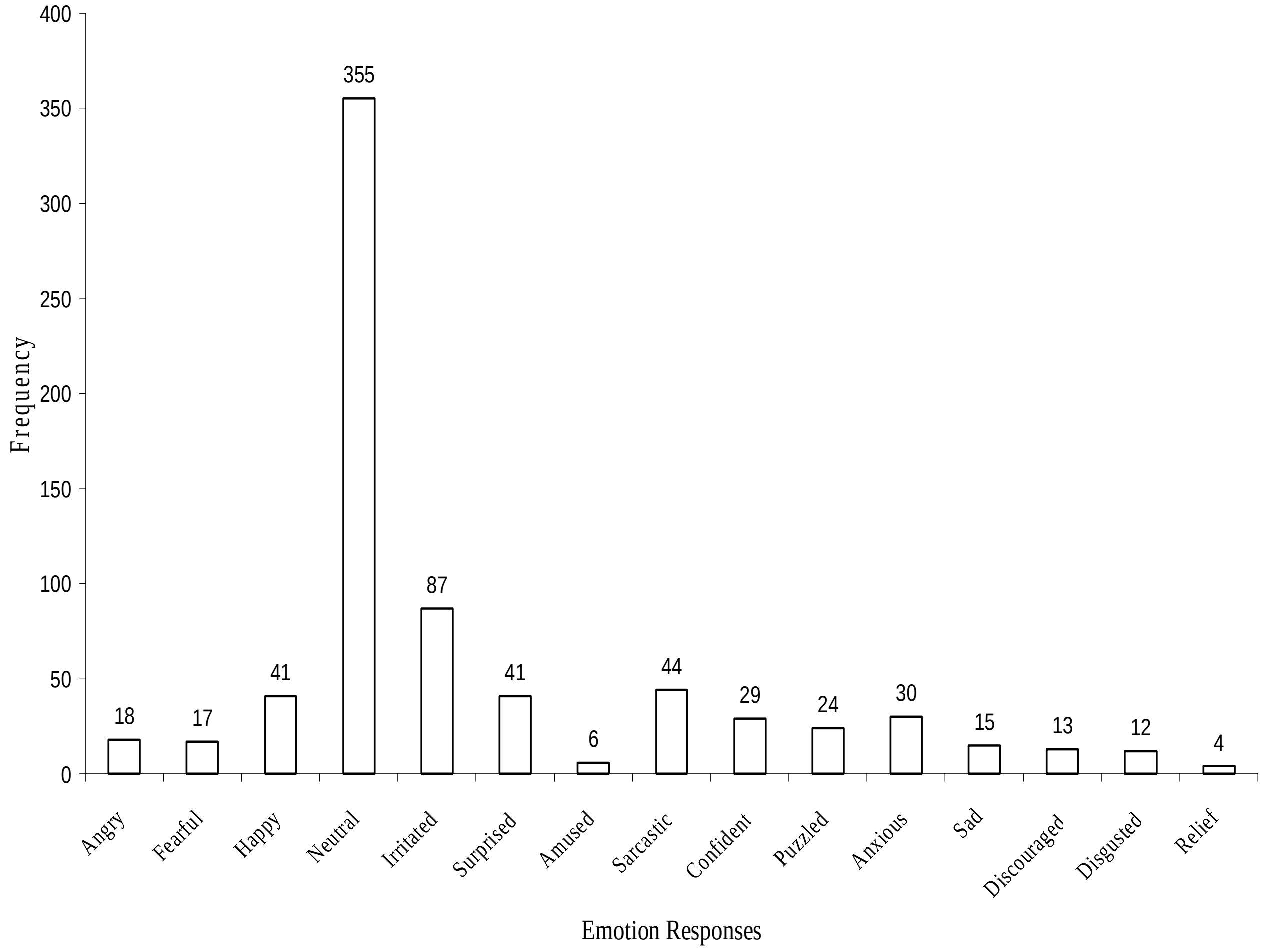

Figure 22 Frequency of Responses for Neutral Face / Happy Voice Condition.

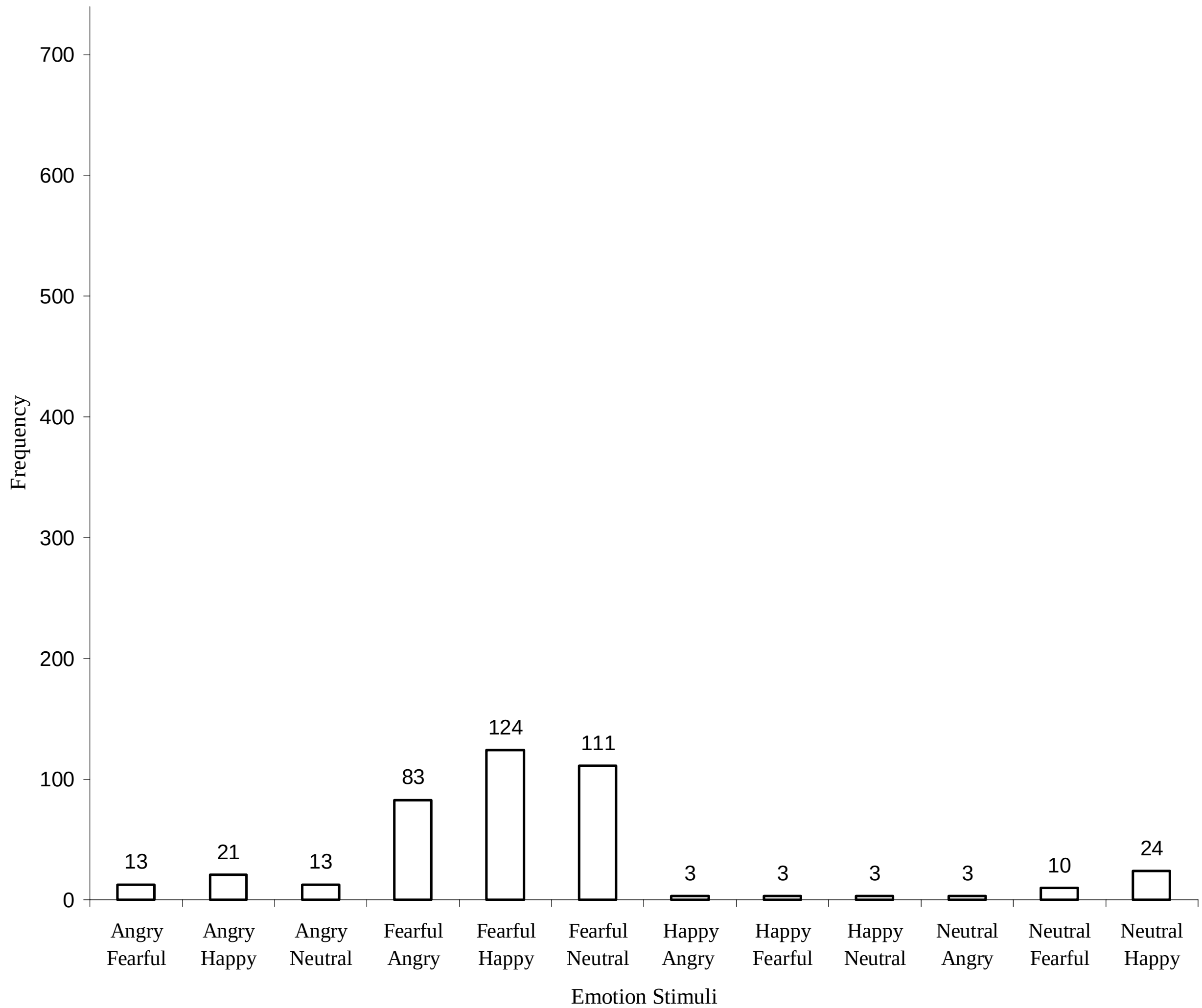

Figure 23 Frequency of Puzzled Responses for Incongruent Stimulus Conditions.

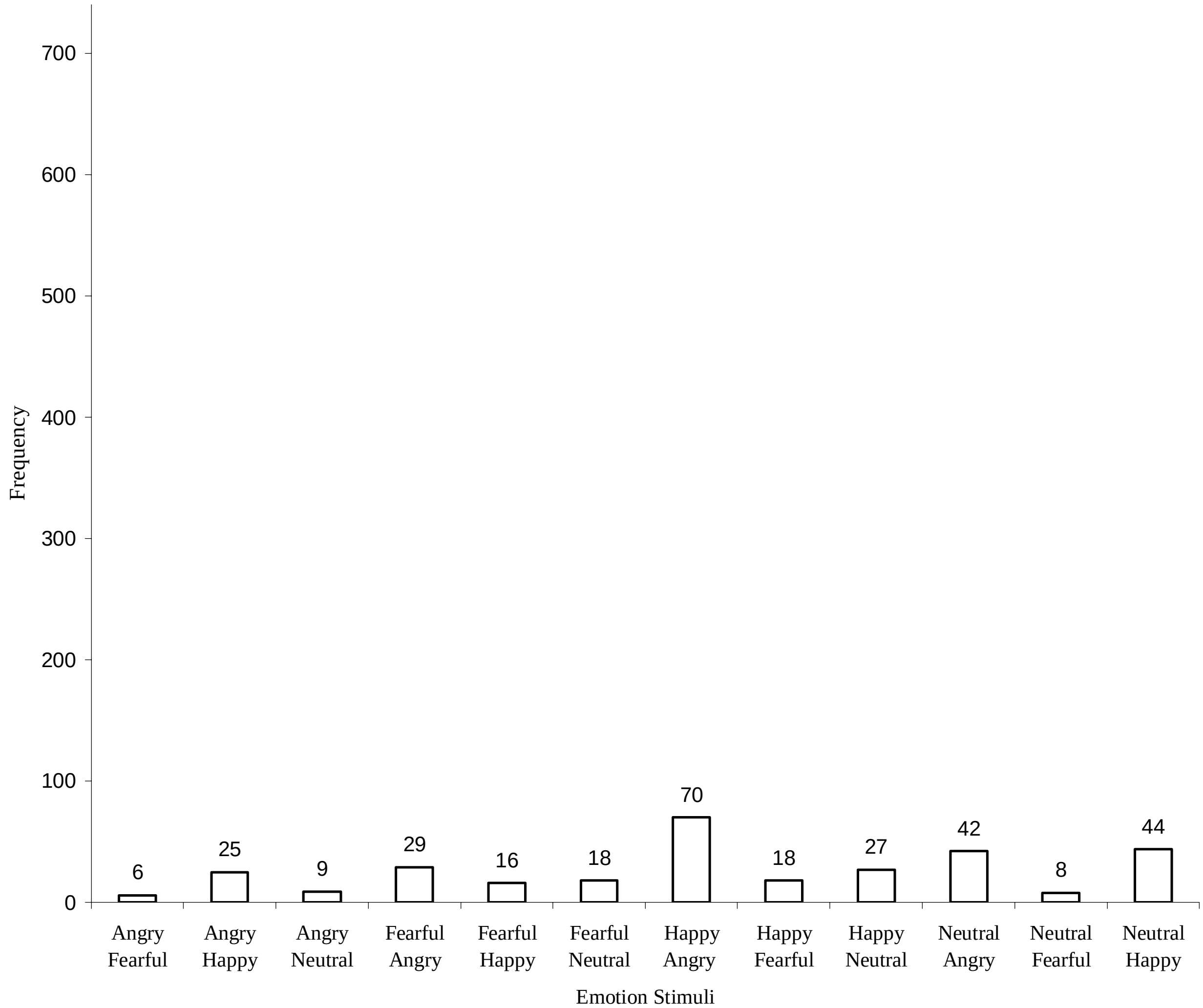

Figure 24 Frequency of Sarcastic Responses for Incongruent Stimulus Conditions.

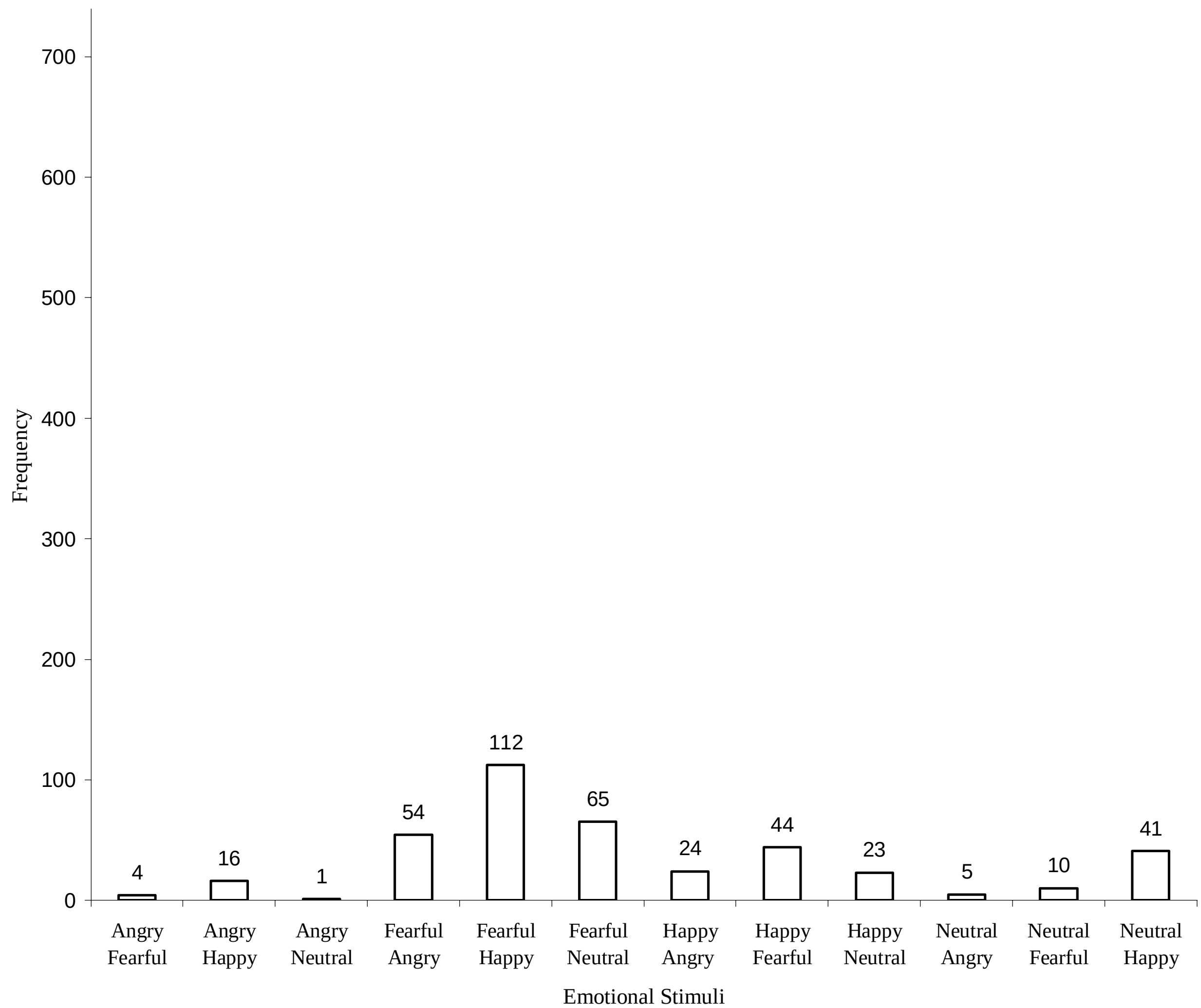

Figure 25 Frequency of Surprised Responses for Incongruent Stimulus Conditions.

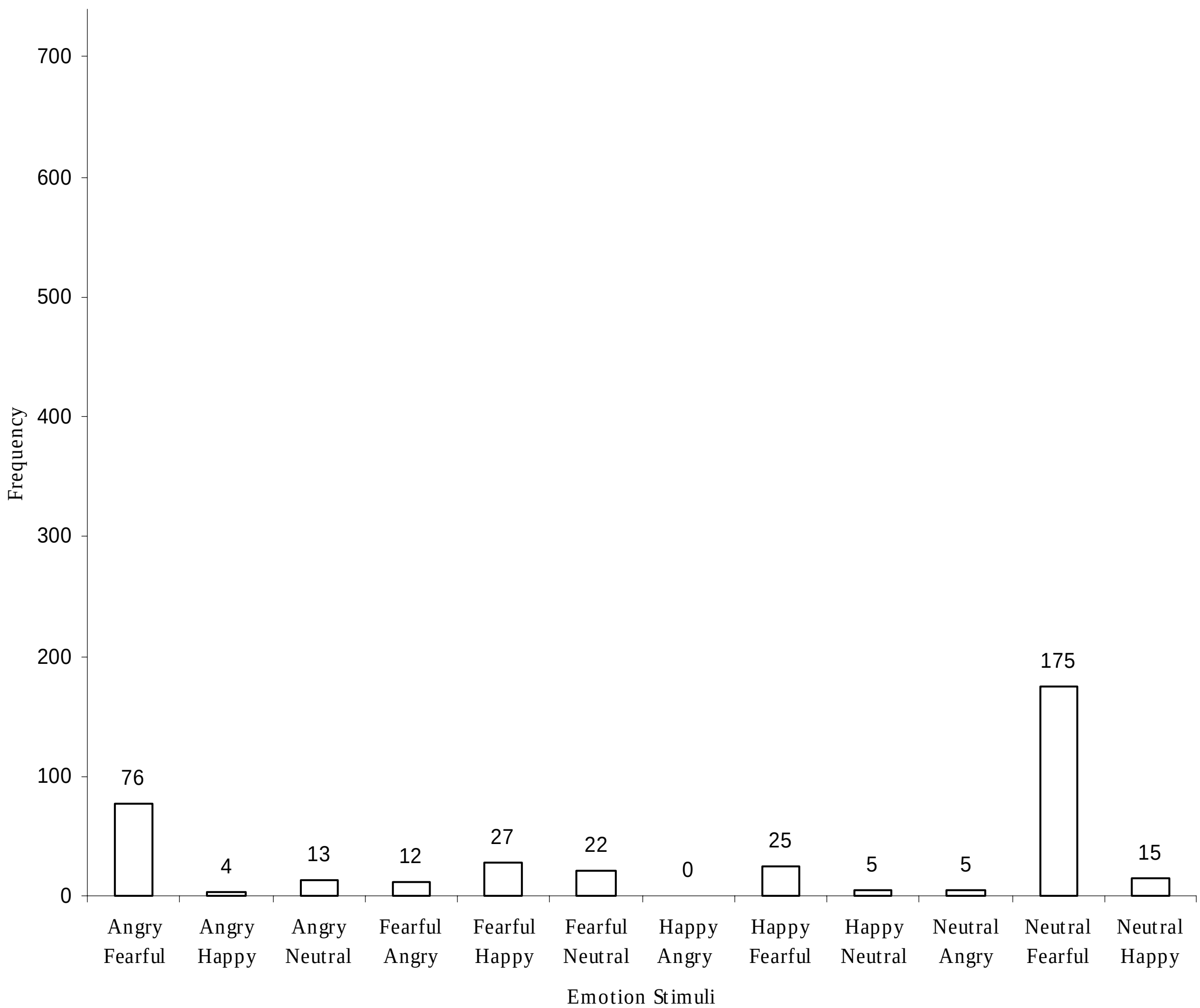

Figure 26 Frequency of Sad Responses for Incongruent Stimulus Conditions.

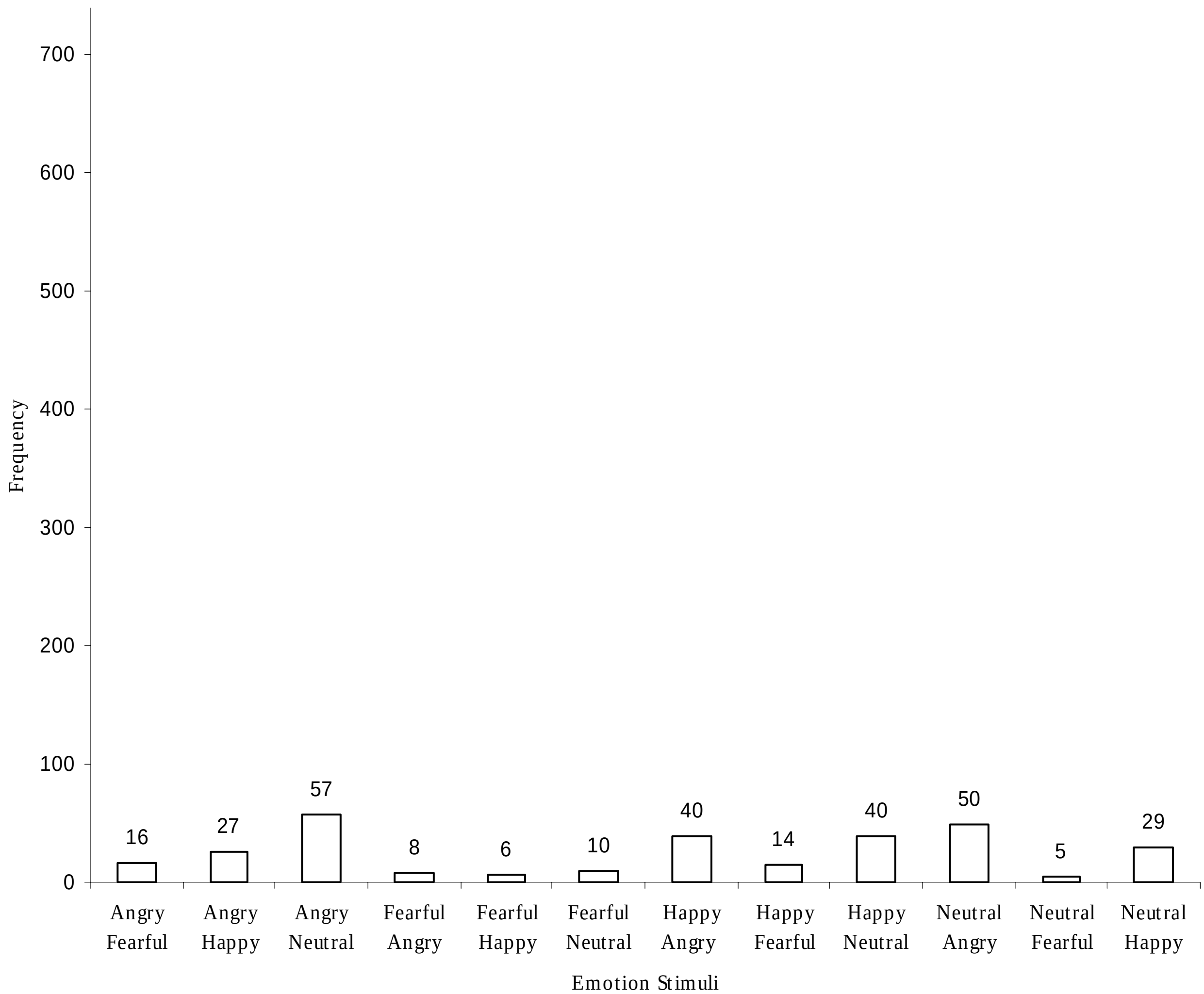

Figure 27 Frequency of Confident Responses for Incongruent Stimulus Conditions.

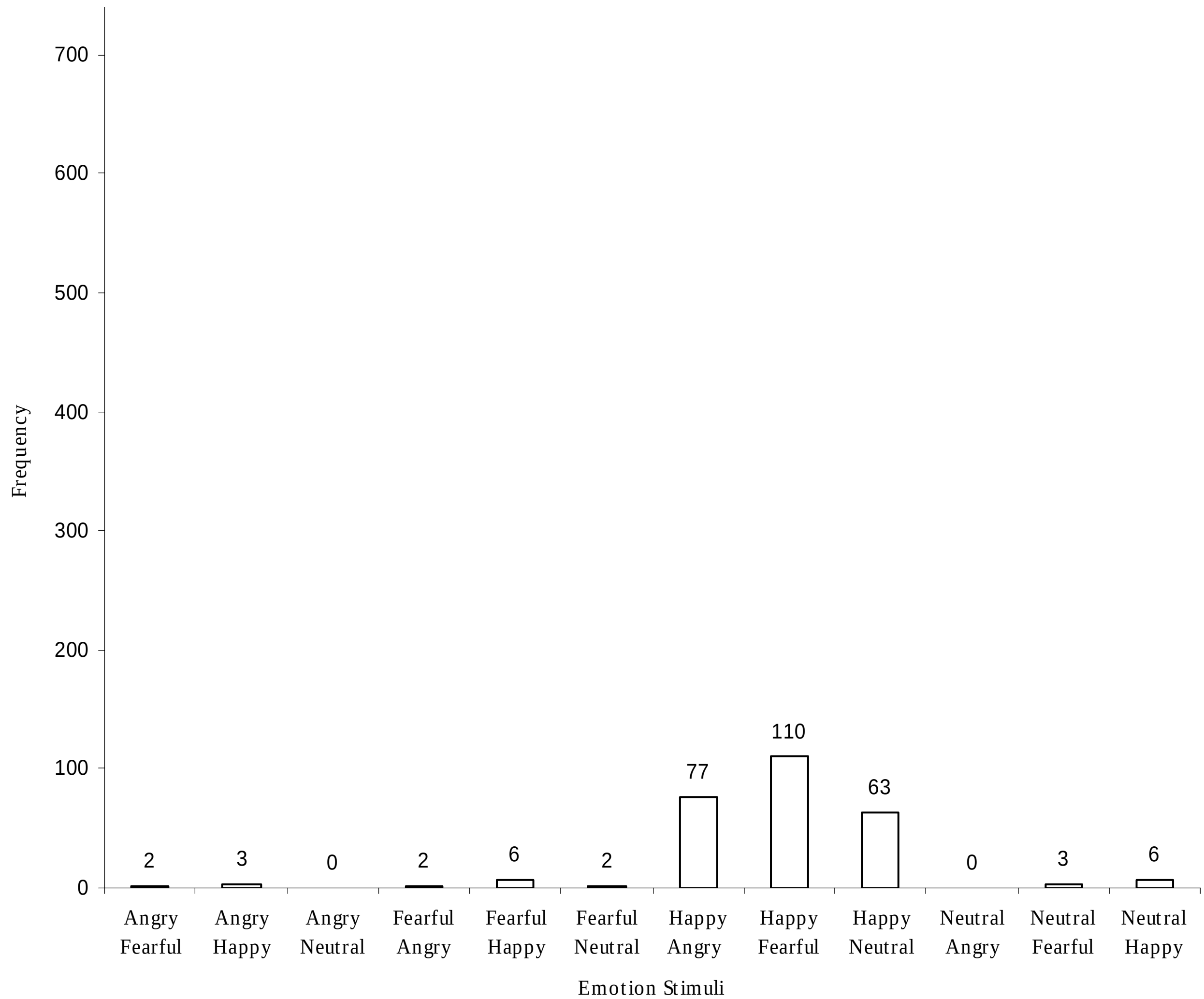

Figure 28 Frequency of Amused Responses for Incongruent Stimulus Conditions.

Discussion

Real-world perception of emotion results from the integration of multiple cues, most notably facial expression and affective prosody. Little research, however, has examined the bimodal perception of emotions. The few studies that have examined bimodal emotion perception have utilized disconnected audio and visual stimuli, resulting in poor ecological validity. The present study investigated the integration of emotional cues from auditory and visual modalities and attempted to generate more realistic emotion processing by using short movies instead of static photographs and disconnected audio recordings. Thirty-seven participants ranging in age from 18-34 were asked to identify the emotions portrayed in dynamic audiovisual stimuli (Robins & Schultz, 2004). Eighty stimuli contained congruent emotions (both face and voice were angry, fearful, happy, or neutral) and 240 contained incongruent emotional cues (e.g., happy face, angry voice). It was hypothesized that for some incongruent conditions one emotion would be more salient, but for other conditions, a blending of the presented emotions would occur, akin to an "emotional McGurk effect." Hypothesis 1 predicted that participants would demonstrate longer reaction times or latency effects when presented with incongruent emotional stimuli. Results indicated that participants did in fact take longer to respond when presented with incongruent emotional stimuli. Although incongruent emotional stimuli resulted in increased latency, responses still occurred within 1 second. Therefore, the current results do not discount or contradict theories of automatic and mandatory cognitive processing of emotions. Rather, the increased response time for incongruent conditions could be accounting for an automatic integrative or evaluation stage. Continued research is needed regarding the time course for the integration and evaluation stages of auditory and visual cues. In fact, a more conclusive answer to the question on the time course of bimodal perception requires additional evidence from other than strictly behavioral methods (deGelder, 2000). For example, research paradigms utilizing neuroimaging, PET, and event-related potential (ERPs) would offer insight into the neural underpinnings of the integration processes. Early work in this area utilizing eletrophysiological measures suggests that both inputs (auditory and visual) are combined at an early stage (deGelder, Vroomen, & Weiskrants, 1999). Continued research utilizing brain-imaging techniques will allow for a better understanding of the relationship between neural processes and behavioral responses, and for the opportunity to develop a more holistic model of bimodal emotion processing.

Hypothesis 2 predicted that participants would correctly identify the emotions presented in the congruent emotional stimuli. Results indicated that participants were in fact able to correctly identify the emotions presented in the congruent stimuli no less than 82 percent of time. Having a high rate of accuracy for the congruent emotional stimuli suggests that the actors were able to accurately convey the emotions they intended to portray. These results also suggest that the responses participants made for the incongruent emotional stimuli were not due to the misperception of the audio or visual components presented in the stimuli.

Hypothesis 3 predicted that participants would respond with a bias of facial expression when presented with incongruent emotional stimuli. Sign tests indicated that for all but three incongruent conditions, participants demonstrated a significant bias toward visual rather than auditory emotion. Moreover, hypothesis 4 predicted that certain participants may consistently prefer one sensory modality over the other when presented with incongruent emotional stimuli and that differences in modality preference may be seen across sexes. The modal response across individuals, regardless of their sex, was biased toward the emotion presented by the face, although 2 participants did respond with a modal response reflective of the emotion presented in the voice and 3 participants responded with a modal response reflective of an emotion other than that presented in the face or voice.

Despite the fact that the majority of participants responded with a facial bias across conditions, it is important to note that the relatively small sample size of the current study may not have clearly identified the extent of individual differences in modality preference. Thus, with a larger sample size future studies may see more participant’s responding with an auditory bias or a bias towards emotions other than those presented in the stimuli. In addition, future research on specific individual predictors such as personality, positive and negative affect, and attributional style is warranted as these specific characteristics may be predictive of modality preferences or biases. However, participants’ preferential attending to facial expressions in the current study supports the notion that facial expressions serve important communicative functions.

The biological perspective of emotion processing holds that emotions are essentially tools designed to regulate behavior in relation to biological evolution. Thus, emotion pervaded the critical ecological problems that our distant ancestors had to solve if their genes were to be represented in the next generation. These problems included finding and sustaining food and drink, finding shelters, asserting oneself socially, engaging with sexual partners, and escaping potentially life-threatening situations. It can be argued that all of these activities are structured by underlying emotions (see Tooby & Cosmides, 1990). Therefore, from a biological perspective, emotions can be understood as being shaped by natural selection and evolution (Öhman, Flykt, & Lundqvist, 2000).

Moreover, it is generally accepted that facial expressions represent innate and automatic behavior patterns determined by evolutionary selection (Darwin, 1872). In fact, studies have identified a number of emotions that can be recognized across both culture and race (Ekman, 1992) and developmental studies indicate an innate predisposition to mimic and discriminate between facial expressions (Field, Woodson, Greenber, & Cohen, 1982; Meltzoff & Moore, 1977). Psychophysiological studies of facial mimicry in adults suggest that processing of emotional expressions is obligatory and largely independent of voluntary processes (Öhman & Dimberg, 1978).

Based on previous findings regarding the importance of facial expressions in biological preparedness it is not surprising that the majority of incongruent conditions in the current study elicited responses reflective of the facial expression presented. However, despite the current findings suggesting the salience of facial expressions in bimodal emotion perception, further research needs to be conducted to determine the specific roles of prosody in both biological preparedness and as a communicative agent. Although facial expressions have been found to be consistent across cultures and have thus argued for a universal and evolutionarily adaptive function, when studied apart from facial expressions prosody may produce similar findings. Therefore, the current findings may suggest that in these specific incongruent conditions facial expressions were more salient than prosody, but may be underestimating the role of prosody in evolution and across cultures. Future research should gain a better understanding of prosody in these roles before underestimating its importance in human communication and emotion processing.

Biological preparedness may also explain why individuals perceive certain emotions as more salient than others. For example, Angry was the most frequent response overall and Irritated was the most frequent response after Fearful, Happy, and Neutral responses (all of which were emotions presented in the stimuli). These results may indicate that human beings are biologically prepared to recognize anger, which could indicate a harm or threat to their well-being or safety. Gosselin and colleagues (1997) found that perception sensitivity of certain signals, such as anger, could be advantageous. In the case of anger, perception at a very low level is likely to give a fitness advantage. In fact, stimuli directly related to survival can be described as having innate perceptual salience, determined by evolutionarily selected value systems in the brain (Edelman, 1987; Fr, Tononi, Reeke, Sporns, & Edelman, 1994). An example of a model that provides a biological account of the acquisition of salience is that of Friston et al. (1994). This model assumes that organisms have evolutionarily selected regulatory systems. When activated, these systems signal behavior that would be of value to the organism.

The dominance of the auditory cue in the Neutral Face / Angry Voice condition and the Neutral Face / Fearful voice condition are more surprising. However, these results make sense when considering the fact that angry or fearful emotions may inherently be more salient than neutral presentations. The voice is likely to be more suited to the expressive and communicative needs of certain emotions than of others. For example, there may be a clear adaptive advantage to being able to warn (fear) or threaten (anger) others in a fairly indirect way over large distances (Johnston & Scherer, 2000). However, the combination of a Neutral Face / Happy Voice produced more responses reflective of the facial expression rather than auditory expression. Interestingly, research shows that there seem to be asymmetries between perception of facial and vocal emotions that are contrary to the common-sense view that emotions are equally recognizable despite the modality in which they are presented. For example, previous research has shown that happiness is the easiest emotion to recognize through facial expressions, but when expressed in the voice only, happiness is quite hard to distinguish from a neutral expression (Scherer, 1979; Vroomen, Collier, & Mozziconacci, 1993). Thus, the current results are consistent with previous literature that has found happy expressions in the voice are hard to distinguish from a neutral expression. It could also be argued that it may be more adaptive for an individual to be able to detect fear or anger through vocal expressions than it would be to detect happiness (Johnstone & Scherer, 2000).

Despite the salience of facial expressions across conditions, ancillary analyses indicated that when participants chose a response other than the emotion expressed in the face, they were still likely to choose a response congruent with the general affective tone portrayed by the facial expression. Therefore, facial expressions not only marked specific emotions, but also established the general affective tone of the stimulus. Clinically, it may be important to examine the role that a therapist's nonverbal cues, especially facial expressions, play in establishing the general affective tone in a therapy situation. Additionally, clinicians' emphasis on verbal rather than nonverbal behaviors (particularly facial expressions) may overlook core aspects of the development, maintenance and treatment of affective disorders.

Despite the dominance of the facial expression on participant responses, as predicted by hypothesis 5, certain conditions elicited responses that were not consistent with the emotion portrayed in the face or voice. These responses may be evidence in support of an "emotional McGurk" effect or Plutchik's explanation of secondary emotions. Plutchik's psychoevolutionary theory of emotion (see, e.g., Plutchik, 1980; 1984) argues that "some emotions are fundamental, or primary, and others are derived or secondary, in the same sense that some colors are primary and others are mixed" (Plutchik, 1984, p. 200). According to Plutchik (1984), complex emotions, such as love or contempt, may be seen as combinations of basic emotions. Love, for example, according to Plutchik, is a combination of joy and acceptance.

The current study found results that can be thought of as consistent with an “emotional McGurk effect” or Plutchik’s psychoevolutionary theory of emotion. Certain combinations of emotions resulted in a higher frequency of responses that reflected a response other than that presented in the face or the voice. For example, the current study found that the presentation of specific combinations of incongruent stimuli resulted in specific responses that were different from the emotions presented in the face or voice (e.g. Happy Face with Angry Voice → Sarcastic; Fearful Face with Happy Voice → Surprised; Neutral Face with Fearful Voice → Sad; etc.). Future research should further explore specific combinations of emotions that tend to elicit derived or secondary emotions. The further exploration of what can be thought of as “blended” emotions may offer a compromise between evolutionary and constructivist viewpoints on the function of emotions. Certain emotions may be biologically determined while other emotions (secondary emotions) may exist because of socially constructed reactions to the presentation of more than one emotion.

Additionally, when presented with incongruent emotional stimuli, participants who did not select the emotion portrayed by either the face or the voice tended to choose a third emotion that was congruent with the valence of the affective tone portrayed by the facial expression. Therefore, facial expressions not only marked specific emotions, but also established the general affective tone of the stimulus. Clinically, it may be important to examine the role that therapist nonverbal cues, especially facial expressions, play in establishing the general affective tone in a therapy situation. Moreover, clinicians’ emphasis on verbal rather than nonverbal behaviors (particularly facial expressions) may overlook core aspects of the development, maintenance and treatment of affective disorders.

Limitations and Future Research

The current sample included an unequal number of males and females. A sample that includes equal numbers of male and female participants may be better able to accurately identify sex differences that may occur across response patterns. In addition, the current study did not assess for participants' racial identity. Group differences in response patterns may be seen across different racial or ethnic groups. Similarly, personality traits and affective states may significantly contribute to individual perceptions of emotions. Future research should further investigate the role of sex differences, racial backgrounds, individual personality traits and affective states in the bimodal perception of emotion. Taking these factors into account will improve the generalizability of findings.

The current study utilized a forced-choice format to assess participant responses to the emotional stimuli. However, this method may have influenced participant responses. For example, participants may have perceived an emotion that was not listed as a choice or may have felt compelled to select all emotion choices at least once. Moving to a free-response format in future research could more accurately assess the bimodal perception of incongruent emotions. Moreover, the current response choices may not have been the most representative of basic, mild forms, and more complex emotions. Extant literature varies in the explanation and selection of basic emotions. For example, whereas the current study found surprise to be a more complex emotion likely resulting from the combination of a Fearful Face with a Happy Voice, Ekman, Izard, and Plutchik argue that surprise is in fact a basic emotion. Still others (e.g. McDougall, Mowrer, and Panksepp) do not include surprise as a basic emotion (see Ortony & Turner, 1990). Future research should attempt to utilize emotions from one theory if using a forced-choice format or make an effort to directly compare theories. Additionally, future research should attempt to further explore the strengths and weaknesses of the existing theories of basic and complex emotions.

Future research of bimodal emotion perception may employ a paradigm similar to that employed by deGelder & Vroomen (2000). Their paradigm instructed participants to ignore part of the stimulus in an attempt to examine the automaticity of bimodal emotion processing. For example, the participant would be instructed to make their decisions about the emotion being portrayed based on solely one modality and to ignore the information they were receiving from the other, incongruent, modality. The Stroop paradigm (Stroop, 1935) is perhaps the most well known illustration of a way in which multiple sources of information can be integrated. In Stroop tasks, some aspect of a stimulus is either irrelevant or incompatible with completion of the primary task. Performing well on such tasks requires the maintenance of attentional focus and executive control. These tasks prove to be difficult for most individuals because they require the inhibition of automatic responses and thus require the individual to control or ignore a very salient but irrelevant source of information (Wurm, Labouvie-Vief, Aycock, Rebucal, & Koch, 2004). Previous research has utilized the Stoop task as an emotion-related task by using stimuli that have some relation to mood, emotion processing, or emotion regulation. Such “emotional Stroop” tasks are widely used to examine complex integrative processes involving the interaction of cognition and emotion. The paradigm utilized by deGelder & Vroomen (2000) may be akin to an “emotional Stroop” rather than an “emotional McGurk” effect. Attentional components involved in emotion processing may speak to the automatic nature of integration in bimodal emotion perception. If the integration of auditory and visual components in emotion processing are automatic, it would be difficult for a participant to successfully ignore one of these modalities despite instructions to do so. Results from deGelder & Vroomen (2000) suggest that participants were unable to focus on only one modality and thus may be evidence for the automatic nature of bimodal integration of emotional cues. deGelder & Vroomen (2000) suggest that this may be an example of an “emotional McGurk effect.” The current findings suggest that while responses tended to be reflective of the emotion presented in the face, certain conditions did result in a response reflective of the integration of the two modalities. However, at this point research can not determine whether or not this effect was due to an “emotional McGurk” effect or an “emotional Stroop.” Thus, future research should examine whether individuals are able to deliberately control the deployment of their attention to auditory versus visual components of incongruent emotions or whether these shifts in attention truly occur automatically. Future research should examine time differences between paradigms in which they are told to ignore one modality and instances when they are not given specific instructions regarding their response choices.

Future research should also investigate the impact of different developmental stages in bimodal emotion perception. The investigation of the myriad ways in which children learn to perceive and attend to emotions will likely provide a more complete picture of the complex interactions involved in bimodal emotion perception. Although the current study found a modality preference for visual stimuli, prior reports suggest that modality preferences change with development (Robinson & Sloutsky, 2004). The investigation of emotion perception throughout different developmental stages will allow for a better understanding of adult emotion perception. Gaining a better understanding of the normal development of emotion perception may also allow for a greater understanding of deficits in emotion perception. Abnormal or deficient emotion processing has been considered to be a hallmark of psychopathology. Despite this link, the underlying relationship between psychopathology and emotion processing remain obscure. Moreover, the relation between specific forms of psychopathology and the processing of emotions are unclear. A greater understanding of the salience of certain emotions for populations with psychopathology may help to inform both development and treatment of these individuals. Future research should consider utilizing dynamic AV stimuli in longitudinal designs of children across different developmental stages, and in studies examining different forms of psychopathology. Utilizing dynamic emotional stimuli may provide more ecologically valid understandings of emotion processing in these populations.

The current study examined bimodal emotion perception through the integration of emotional cues from both auditory and visual modalities. The dynamic emotional stimuli used in this study attempted to generate more realistic understandings of emotion processing, similar to that encountered in real world situations. Matching emotion conditions elicited a significant number of correct responses for all four emotions, while mismatching emotion conditions tended to result in a bias towards the visual modality. Despite these general findings, certain mismatching conditions resulted in specific “blends” of the emotions presented in the face and voice. The present study provides only a first examination of bimodal emotion perception through the use of both dynamic and incongruent emotional stimuli. As such, it raises many new questions and possibilities. Answers to these questions can help significantly advance knowledge of emotion processing in various populations, the automatic nature of integration, and the neural underpinning of these processes.

References

Adolphs, R., Tranel, D., Damasio, H., & Damasio, A. (1994). Impaired recognition of emotion in facial expressions following bilateral damage to the human amygdale. *Nature, 372* (6507), 669-672.

Adolphs, R. (2001). The neurobiology of social cognition. *Current Opinion in Neurobiology, 11*, 231-239.

Adolphs, R., Sears, L., & Piven, J. (2001). Abnormal processing of social information from faces in autism. *Journal of Cognitive Neuroscience, 13* (2), 232-240.

Aggleton, J. P. (1992). The functional effects of amygdale lesions in humans: A comparison with findings from monkeys. In J. P. Aggleton (Ed.), *The amygdale: Neurobiological aspects of emotion, memory and mental dysfunction.* (pp. 485-503). New York: Wiley-Liss.

Amaral, D.G., Bauman, M.D., Schumann, C.M. (2003). The amygdala and autism: Implications from non-human primate studies. *Genes, Brain & Behavior, 2* (5), 295-302.

Ambady, N., & Gray, H.M. (2002). On being sad and mistaken: Mood effects on the accuracy of thin-slice judgments. *Journal of Personality and Social Psychology, 83,* 947-961.

American Psychiatric Association, (1994). *Diagnostic and statistical manual of mental disorders* (4^{th} ed.). Washington, DC.

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

Baron-Cohen, S., Ring, H., Wheelwright, S., Bullmore, E., Brammer, M., Simmons, A., & Williams, S. (1999). Social intelligence in the normal and autistic brain: an fMRI study. *European Journal of Neuroscience, 11,* 1891-1898.

Baron-Cohen, S., Ring, H., Bullmore, E., Wheelwright, S., Ashwin, C., & Williams, S. (2000). The Amygdale theory of autism. *Neuroscience and Biobehavioral Reviews, 24*(3), 355-364.

Baum, K. M., & Nowicki, S. Jr. (1998). Perception of emotion: Measuring decoding accuracy of adult prosodic cues varying in intensity. *Journal of Nonverbal Behavior, 22,* 89-107.

Bennett, D.S., Bendersky, M., & Lewis, M. (2005). Antecedents of emotion knowledge: Predictors of individual differences in young children. *Cognition and Emotion, 19* (3), 375-396.

Benton, A.L. (1994). *Face Recognition.* Los Angeles: Western Psychological Services.

Blair, R.J., Morris, J.S., Frith, C.D., Perrett, D.I., Dolan, R.J. (1999). Dissociable neural responses to facial expressions of sadness and anger. *Brain, 122,* 883-893.

Boucher, J., Lewis, V., & Collis, G. (2000). Voice processing abilities in children with autism, children with specific language impairments, and young typically developing children. *Journal of Child Psychology and Psychiatry, 41* (7),847-857.

Breiter, H.C., Etocff, N.L., Whalen, P.J., Kennedy, W.A., Rauch, S.L., Buckner, R.L., Strauss, M.M., Hymann, S.E., Rosen, B.R. (1996). Response and habituation of the human amygdale during visual processing of facial expression. *Neuron* (17), 875-887.

Brody, L.R. (1985). Gender differences in emotional development: A review of theories and research. *Journal of Personality, 53,* 102–149.

Bruno, N., & Cutting, J. E. (1988). Minimodularity and the perception of layout. *Journal of Experimental Psychology: General, 117,* 161-170.

Bushara, K. O., Grafman, J., & Hallett, M. (2001). Neural correlates of auditory-visual stimulus onset asynchrony detection. *Journal of Neuroscience, 21,* 300-304.

Calvert, G. A. (2001). Crossmodal processing in the human brain: Insights from functional neuroimaging studies. *Cerebral Cortex*, *11*, 1110-1123.

Calvert, G. A., Hansen, P. C., Iversen, S. D., & Brammer, M. J. (2001). Detection of audio visual integration sites in humans by application of electropsysiological criteria to the BOLD effect. *NeuroImage, 14*, 427-438.

Celani, G., Battacchi, M.W., & Arcidiacono, L. (1999). The understanding of the emotional meaning of facial expressions in people with autism. *Journal of Autism and Developmental Disorders, 29* (1), 57-66.

Cohen, J. et al. (1994). *PsyScope*. Carnegie Mellon University, Department of Psychology.

Cornelius, R. R. (1996). *The science of emotion: Research and tradition in the psychology of emotion*. New Jersey: Prentice-Hall, Inc.

Crick, N.R., & Dodge, K.A. (1994). A Review and reformulation of social information-processing mechanisms in children's social adjustment. *Psychological Bulletin, 115,* 74-101.

Cutting, J.E., Bruno, N., Brady, N. P., & Moore, C. (1992). Selectivity, Scope, and Simplicity of Models: A lesson from fitting judgments of perceived depth. *Journal of Experimental Psychology: General*, *121* (3), 364-381.

Damasio, A.R. (1989). Time-locked multiregional retroactivation: A systems-level proposal for the neural substrates of recall and recognition. *Cognition, 33*, 25-62.

Damasio, A.R., Grabowski, T.J., Bechara, A., Damasio, H., Ponto, L.L., Parvizi, J., Hichwa, R.D. (2000). Subcortical and cortical brain activity during the feeling of self-generated emotions. *Nature Neuroscience, 3,* 1049-1056.

Darwin, C. (1872 [1965]). *The Expression of the Emotions in Man and Animals.* Chicago: University of Chicago Press.

Dawson, G. (1996). Neuropsychology of autism: a report on the state of the science. *Journal of Autism and Developmental Disorders, 26,* 179-184.

Dawson, G., Meltzoff, A.N., Osterling, J., & Rinaldi, J. (1998). Neuropsychological correlates of early symptoms of autism. *Child Development, 69* (5), 1276-1285.

deGelder, B. (2000). Recognizing emotions by ear and by eye. Lane, R. D., & Nadel, L. (Eds.), *Cognitive Neuroscience of Emotion* (pp. 84-105). New York: Oxford University Press.

deGelder, B., Vroomen, J., & Weiskrantz, L. (1999). Covert processing of facial expressions in a blindsight patient. Annual Meeting of the Cognitive Neuroscience Society, Abstract 35A, Washington, D.C. In Lane, R.D., & Nadel, L. (Eds.), *Cognitive Neuroscience of Emotion*. New York: Oxford University Press.

deGelder, B., & Vroomen, J. (2000). The perception of emotions by ear and by eye. *Cognition and Emotion, 14*(3), 289-311.

de Gelder, B., & Bertelson, P. (2003). Multisensory integration, perception and ecological validity. *Trends in Cognitive Sciences, 7*(10), 460-467.

de Gelder, B., Vroomen, J., Annen, L., Masthof, E., & Hodiamont, P. (2003). Audio-visual integration in schizophrenia. *Schizophrenia Research, 59*(2-3), 211-218.

de Gelder, B., Vroomen, J. & Pourtois, G. (2004). Multisensory perception of affect, its time course and its neural basis. In Calvert, G., Spence, C. & Stein, B.E. (Ed.), *Handbook of multisensory processes.* (pp. 581-596). Cambridge, MA: MIT.

de Gelder, B., Vroomen, J., de Jong, S.J., Masthoff, E.D., Trompenaars, F.J., Hodiamont, P. (2005). Multisensory integration of emotional faces and voices in schizophrenics. *Schizophrenia Research, 72*, 195-203.

De Silva, L.C., Miyasato, T., & Nakatsu, R. (1997). *Facial Emotion Recognition Using Multimodal Information.* In Proc. IEEE Int. Conf. on Information, Communications, and Signal Processing (ICICS'97), Singapore, pp. 397-401.

Edelman, G. (1987). *The Remembered Present: A Biological Theory of Consciousness.* New York: Basic Books.

Emery, N. J., Capitanio, J. P., Mason, W. A., Machado, C. J., Mendoza, S. P., Amaral, D. G. (2001). The Effects of bilateral lesions of the amygdala on dyadic social interactions in rhesus monkeys (Macaca mulatta). *Behavioral Neuroscience, 115*, 515-544.

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

Ekman, P. & Friesen, W.V. (1976). Measuring facial movement. *Environmental Psychology & Nonverbal Behavior, 1* (1), 56-75.

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

Ekman, P. & Heider, K. G. (1988). The universality of a contempt expression: A replication. *Motivation and Emotion, 12*, 303-308.

Esslen, M., Pascual-Marqui, R.D., Hell, D., Kochi, K., & Lehmann, D. (2004). Brain areas and time course of emotional processing. *NeuroImage, 21,* 1189-1203.

Ettlinger, G., & Wilson, W. A. (1990). Cross-modal performance: Behavioral processes, phylogenetic considerations and neural mechanisms. *Behavioral and Brain Research, 40,* 169-192.

Field, T., Woodson, R., Greenber, R., & Cohen, D. (1982). Discrimination and imitation of facial expressions in neonates. *Science, 218,* 179-181.

Fridlund, A. J. (1992). The behavioral ecology and sociality of human faces. M.S. Clark (Ed.), Emotion: Review of Personality and Social Psychology, Vol. 13 (pp. 90-121). Newbury Park, CA: Sage.

Fridlund, A. J. (1994). Human facial expression: An evolutionary view. San Diego: Academic Press.

Fridlund, A., Ekman, P. & Oster, H. (1987). Facial expressions of emotion. In A. Siegman & S. Feldstein (Eds.), *Nonverbal behavior and communication,* (pp. 143-224). Hillsdale, NJ: Erlbaum.

Friston, K. J., Tononi, G., Reeke, G. N., Sporns, O., & Edelman, G. M. (1994). Value-dependent selection in the brain: simulation in a synthetic neural model. *Neuroscience, 30,* 77-86.

Gadow, K.D., Sprafkin, J., & Weiss, M.D. (1999). *Adult Self-Report Inventory-4 Manual* Stony Brook, NY: Checkmate Plus.

Gauthier, I., Tarr, M. J., Anderson, A. W. (1999). Activation of the middle fusiform “face area” increases with expertise in recognizing novel objects. *Nature Neuroscience, 2* (6), 568-573.

Gauthier, I., Skudlarski, P., Gore, J. C., Anderson, A.W. (2000). Expertise for cars and birds recruits brain areas involved in face recognition. *Nature Neuroscience, 3* (2), 191-197.

Gepner, B., de Schonen, S., & Buttin, C. (1994). Face processing in young autistic children. *Infant Behavior and Development, 17,* 661.

Gosselin, P., Kirouac, G., Dore, F.Y. (1997). Components and recognition of facial expression in the communication of emotion by actors. Ekman, P., Rosenberg, E. (Eds). *What the face reveals* (pp. 169-222). New York: Oxford University Press.

Green, K.P. & Gerdeman, A. (1995). Cross-Modal discrepancies in coarticulation and the integration of speech information: the McGurk effect with mismatched vowels. *Journal of Experiment Psychology: Human Perception and Performance.* Vol 21(6), 1409-1426.

Herba, C., & Phillips, M. (2004). Annotation: Development of facial expression recognition from childhood to adolescence: behavioral and neurological perspectives. *Journal of Child Psychology and Psychiatry, 45,* 1185-1198.

Hill, E., Berthoz, S., & Frith, U. (2004). Brief report: Cognitive processing of own emotions in individuals with autistic spectrum disorder and in their relatives. *Journal of Autism and Developmental Disorders*, *34*, 229-235.

Hobson, R.P., Ouston, J., & Lee, A. (1988). What's in a face? The case of autism. *British Journal of Psychology*, *79*(4), 441-453.

Huber, L., & Lenz, R. (1993). A test of linear feature model of polymorphous concept discrimination with pigeons. *Quarterly Journal of Experimental Psychology, 46 B,* 1-18.

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

Izard, C.E., & Harris, P. (1995). Emotional development and developmental psychopathology. In D. Cicchetti & D. J. Cohen (Eds.), *Developmental psychopathology: Vol. 1: Theory and methods* (pp. 467-503). New York: Wiley.

Johnson, W. F., Emde, R., Scherer, K., & Klinnert, M. D. (1986). Recognition of emotion from vocal cues. *Archives of General Psychiatry, 43,* 280-283.

Johnston, T. & Scherer, K.R. (2000). Vocal communication of emotion. In Lewis, M. & Haviland-Jones, J.M (Eds.), *Handbook of Emotions* (pp. 220-235). New York: The Guilford Press.

Jones, J.A., & Callan, D.E. (2003). Brain activity during audiovisual speech perception: An fMRI study of the McGurk effect. *Cognitive Neuroscience and Neuropsychology, 14* (8), 1129-1133.

Kimbrell, T.A., George, M.S., Parekh, P.I., Ketter, T.A., Podell, D.M., Danielson, A.L., Repella, J.D., Benson, B.E., Willis, M.W., Herscovitch, P., Post, R.M. (1999). Regional brain activity during transient self-induced anxiety and anger in healthy adults. *Biological Psychiatry, 46,* 454-465.

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

MacDonald, J. & McGurk, H. (1978). Visual influences on speech perception processes. *Perception and Psychophysics, 24,* 253-257.

MacDonald, H., Rutter, M., Howlin, P., Rios, P., Leconteur, A., Evered, C., & Folstein, S. (1989). Recognition and expression of emotional cues by autistic and normal adults. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 30* (6), 865-877.

Massaro, D.W. (1987). Categorical partition: A fuzzy logical model of categorization behavior. In S. Harnad (Ed.), Categorical perception: The groundwork of cognition (pp. 254-283). New York: Cambridge University Press.

Massaro, D.W., & Egan, P.B. (1996). Perceiving affect from the voice and the face. *Psychonomic Bulletin and Review, 3,* 215-221.

Massaro, D.W., & Cohen, M.M. (2000). Fuzzy logical model of bimodal emotion perception: Comment on “The perception of emotions by ear and by eye” by de Gelder and Vroomen. *Cognition & Emotion, 14* (13), 313-320.

McCaffrey, R.J., Duff, K., & Westervelt, H.J. (2000). Practitioner's guide to evaluating change neuropsychological assessment instruments. New York: Kluwer Academic/Plenum Press.

McClure, E.B. (2000). A meta-analytic review of sex differences in facial expression processing and their development in infants, children, and adolescents. *Psychological Bulletin*, *126,* 424-453.

McClure, E.B., & Nowicki, S. Jr. (2001). Associations between social anxiety and nonverbal processing skill in preadolescent boys and girls. *Journal of Nonverbal Behavior, 25,* 3-19.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264,* 746-748.

Meltzoff, A. N. & Moore, M. K. (1977). Imitation of facial and manual gestures by human neonates. *Science, 198,* 74-78.

Mesulam, M. M. (1998). From sensation to cognition. *Brain, 121*, 1013 – 1052.

Morris, J.S., Frith, C.D., Perrett, D.I., Rowland, D., Young, A.W., Calder, A.J., Dolan, R.J. (1996). A differential neural response in the human amygdale to fearful and happy facial expressions. *Nature, 383,* 812-815.

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

Nelson, C. A. (1987). The recognition of facial expressions in the first two years of life: Mechanisms of development. *Child Development, 58* (4), 889-909.

Nowicki, S. (2004). *Instruction manual for the receptive tests of the Diagnostic Analysis of Nonverbal Accuracy 2.* Unpublished manuscript, Emory University, Atlanta, GA.

Nowicki, S., & Carton, J. (1993). The measurement of emotional intensity from facial expressions. *The Journal of Social Psychology, 133,* 749-750.

Nowicki, S., & Duke, M.P. (1994). Individual differences in the nonverbal communication of affect: The Diagnostic Analysis of Nonverbal Accuracy Scale. Special Issue: Development of nonverbal behavior: II. Social development and nonverbal behavior. *Journal of Nonverbal Behavior, 18,* 9-35.

Öhman, A., & Dimberg, U. J. (1978). Facial expressions as conditioned stimuli for electrodermal responses: a case of “preparedness”? *Journal of Personality and Social Psychology, 36*, 1251-1258.

Öhman, A., Flykt, A., Lundqvist, D. (2000). Unconscious emotion: Evolutionary perspective, psychophysiological data and neuropsychological mechanisms. Lane, R. D., Nadel, L. *Cognitive neuroscience of emotion* (pp. 296-327). New York: Oxford University Press.

Olson, I. R., Gatenby, J. C., & Gore, J. C. (2002). A comparison of bound and unbound audio-visual information processing in the human cerebral cortex. *Cognitive Brain Research, 14*, 129-138.

Ortony, A. & Turner, T.J. (1990). What’s basic about basic emotions? *Psychological Review, 97* (3), 315-331.

Phan, K.K., Wager, T., Taylor, S.F., & Liberzon, I. (2002). Review: Functional neuroanatomy of emotion: A meta-analysis of emotion activation studies in PET and fMRI. *Neuroimage, 16*, 331–348.

Phillips, M.L., Young, A.W., Senior, C., Brammer, M., Andrew, C., Calder, A.J., Bullmore, E.T., Perrett, D.I., Rowland, D., Williams, S.C.R., Gray, J.A., David, A.S. (1997). A specific neural substrate for perceiving facial expressions of disgust. *Nature, 389,* 495-498.

Phillips, M.L., Drevets, W.C., Rauch, S.L., & Lane, R. (2003). Neurobiology of emotion perception I: The neural basis of normal emotion perception. *Biological Psychiatry, 54,* 504–514.

Plutchik, R. (1980). Emotion: A psychoevolutionary synthesis. New York: Harper and row.

Plutchik, R. (1984). Emotions: A general psychoevolutionary theory. In K. R. Scherer & P. Ekman (Eds.), Approaches to emotion (pp. 197 – 219). Hillsdale, NJ: Erlbaum.

Pollack, I., Rubenstein, H., & Horowitz, A. (1960). Communication of verbal modes of expression. *Language and Speech, 3,* 121-130.

Pollack, S.D., & Kistler, D.J. (2002). Early experience is associated with the development of categorical representations for facial expressions of emotion. *Proceedings of the National Academy of Sciences of the United States of America, 99,* 9072-9076.

Psychological Corp. (1999). Wechsler Abbreviated Scale of Intelligence manual. San Antonio, TX: Author.

Robins, D.L. & Schultz, R.T. (2004). Bimodal integration of emotional information. Paper presented at the International Neuropsychological Society, February, 2004, Baltimore, MD.

Robinson, C. W., & Sloutsky, V. M. (2004). Auditory dominance and its change in the course of development. *Child Development, 75,* 1387 -1401.

Roelfsema, P. R., Engel, A. K., Konig, P., & Singer, W. (1997). Visuomotor integration is associated with zero time-lag synchronization among cortical areas. *Nature, 385*, 157-161.

Rosenblum, L.D., Schmuckler, M.A., & Johnson, J.A. (1997). The McGurk effect in infants. *Perception and Psychophysics,* 59, 347-357.

Rothman, A.D., & Nowicki, S. (2004). A measure of the ability to identify emotion in children's tone of voice. *Journal of Nonverbal Behavior, 28,* 67-92.

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology, 39*, 1161–1178.

Saldana, H.M. & Rosenblum, L.D. (1993). Visual influences on auditory pluck and bow judgments. *Perception & Psychophysics, 54*, 406-416.

Salovey, P., Bedell, B.T., Detweiler, J.B., & Mayer, J.D. (2000). Current directions in emotional intelligence research. In M. Lewis & J.M. Haviland-Jones (Eds.), *Handbook of emotions* (pp. 504-520). New York: Guilford Press.

Scherer, K.R. (1979). Non-linguistic vocal indicators of emotions and psychopathology. In C. E. Izard (Ed), *Emotions in Personality and Psychopathology*. New York: Plenum Press.

Scherer, K.R. (1999). Universality of emotional expression. In D. Levinson, J.J. Ponzetti, Jr., & P.F. Jorgenson (Eds.), *Encyclopedia of human emotions* (pp. 669-674). New York: Macmillan.

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

Scott, S., Young, A., Calder, A., Hellawell, D., Aggleton, J.P., & Johnson, M. (1997). Impaired auditory recognition of fear and anger following bilateral amygdale lesions. *Nature, 385* (6613), 254-257.

Seashore, C.E., Lewis, D., & Saetveit, D.L. (1960). *Seashore Measures of Musical Talents* (rev. ed.). New York: Psychological Corporation.

Sprengelmeyer, R., Rausch, M., Eysel, U.T., Przuntek, H. (1998). Neural structures associated with recognition of facial expressions of basic emotions. *Proceeding of the Royal Society of London, Series B: Biological Sciences, 265,* 1927-1931.

Stroop, J.R. (1935). Studies of interference in serial verbal reactions. *Journal of*

Experimental Psychology, 18, 643-662.

Tooby, J. & Cosmides, L. (1990). The past explains the present: emotional adaptations and the structure of ancestral environments. *Ethology and Sociobiology, 11,* 375-424.

VanLancker, D., Cornelius, C., & Kreiman, J. (1989). Recognition of emotional prosodic meanings in speech by autistic, schizophrenic, and normal children. *Developmental Neuropsychology, 5* (2-3), 207-226.

Vroomen, J., Collier, R. & Mozziconacci, S. (1993). Duration and intonation in emotional speech. In Proceedings of the Third European Conference on Speech Communication and Technology, Berlin, pp. 577-580.

Walker, E. (1981). Emotion recognition in disturbed and normal children: A research note. *Journal of Child Psychology and Psychiatry, 22*, 263-268.

Walker, A., & Grolnick, W. (1983). Discrimination of vocal expressions by young infants. *Infant Behavior and Development, 6*, 491-498.

Williams, C. E. & Stevens, K. N. (1972). Emotions and speech: Some acoustic correlates. *Journal of the Acoustical Society of America, 52*, 1238-1250.

Wurm, L.H., Labouvie-Vief, G., Aycock, J., Rebucal, K.A., & Koch, H.E. (2004). Performance in Auditory and Visual Emotional Stroop Tasks: A comparison of older and younger adults. *Psychology and Aging, 19 (3),* 523-535.

Yirmiya, N., Sigman, M. D., Kasari, C., & Mundy, P. (1992). Empathy and cognition in high-functioning children with autism. *Child Development, 63* (1), 150-160.

Young, A., Hellawell, D., Van De Wal, C., & Johnson, M. (1996). Facial expression processing after amygdalotomy. *Neuropsychologia, 34,* 31-39.

Zadeh, L.A. (1965). Fuzzy sets. *Information and Control, 8,* 338-353.