THESIS

# PROSODIC INFLUENCE IN FACE EMOTION PERCEPTION: EVIDENCE FROM BEHAVIORAL MEASURES AND FUNCTIONAL NEAR-INFRARED SPECTROSCOPY

Submitted by

Katherine M. Becker

Department of Psychology

In partial fulfillment of the requirements

For the Degree of Master of Science

Colorado State University

Fort Collins, Colorado

Summer 2017

Master's Committee:

Advisor: Donald C. Rojas

Lucy J. Troup
Patricia Davies

ABSTRACT

PROSODIC INFLUENCE IN FACE EMOTION PERCEPTION: EVIDENCE FROM BEHAVIORAL MEASURES AND FUNCTIONAL NEAR-INFRARED SPECTROSCOPY

The perception of another person's emotional state is formed by the intersection of simultaneously presented affective vocal and facial information. These two channels are highly effective in communicating emotion as either can do so independently. However, it is unclear how these modalities interact and influence perception when they are integrated. The current study sought to disentangle the roles of each modality by manipulating both the vocal and facial components of emotion perception. Voice stimuli were comprised of nonverbal affective vocalizations produced in either a happy, angry, or neutral prosody. Face images were created from morphed continua, composed to two end-point images, of one happy, and one angry face. These stimuli were presented independently and together to fully dissociate the unimodal and bimodal aspects of affect perception. These stimuli were combined in one hybrid block design paradigm which was used in a behavioral experiment and a functional near-infrared spectroscopy experiment. The results indicated that prosody does effect the perception of affective faces and this can be evidenced in both the behavioral and functional imaging data. Moreover, these data suggest that prosody is differentially represented in the brain in a valence specific way. Together, these findings provide strong support for the crucial role of prosody in affect perception.

# ACKNOWLEDGEMENTS

This work would not have been possible without the support of my thesis committee: Dr. Don Rojas, Dr. Lucy Troup, and Dr. Patricia Davies. Their expertise and insight enabled me to design and execute two research projects which I will present in my master's thesis. This journey contained many stumbling points at which my enthusiasm began to falter, but the continued support of my advisor, committee, and family and friends enabled me to finish this project.

TABLE OF CONTENTS

# CHAPTER 1 – GENERAL INTRODUCTION

Affective cues communicated by the face and voice of another individual are automatically, and effortlessly, integrated by the brain to form whole percepts of emotion. This conceptualization of embodied emotion was first proposed by Darwin (1872), who noted the inherent correspondence between an individual's facial expression and their emotional state. The exact communicative role played by facial expressions has been further refined, and it has been established that all facial expressions can be categorized as belonging to one of six basic emotions (happiness, sadness, anger, surprise, fear, and disgust), each characterized by the unique coordination of multiple facial muscles (Ekman & Friesen, 1976; Bartlett, Viola, Sejnowski, Golomb, Larsen, Hager, & Ekman, 1996). Further, Ekman argued that these emotion categories evolved out of an ecological necessity to rapidly and effectively communicate affiliative or avoidance behaviors in threatening environments (Ekman, 1992; Scherer & Kappas, 1988). These skills are critical to effective communication both within groups of people and in inter-personal interactions.

Affective facial expressions are reflections of innate emotional reactions, accompanied by distinct vocal gestures which contain essential paralinguistic and prosodic information about a speaker's identity and emotional state (Patel, Scherer, Björkner, & Sundberg, 2011; Belin, Bestelmeyer, Latinus, & Watson, 2011; Scherer, Johnstone, & Klasmeyer, 2003; Banse & Scherer, 1996). Prosody is produced via the grouping of suprasegmental features (fundamental frequency, rhythm, and amplitude) to produce unique, emotion specific vocalizations (Belin, Fillion-Bilodeau, & Gosselin, 2008; Grandjean, Sander, Pourtois, Schwartz, Segheir, Scherer, & Vuilleumier, 2005; Patel, et al., 2011). While prosody can be conveyed with or without linguistic cues, vocal affect may be best communicated in the absence of linguistic information, which may

inadvertently incorporate more complicated semantic or lexical representations of the emotion (Belin, Fillion-Bilodeau, & Gosselin, 2008). Similar to the ecological role of affective facial expressions in rapid communication, prosodic nonverbal vocalizations are thought to precede the development of linguistic communication, indicating that prosody may play a foundational role in vocal affect perception (Fitch, 2000).

Collectively, these findings underscore the multimodal nature of the production and perception of emotion. Evidence from functional magnetic resonance imaging (fMRI) studies has shown that these modalities may be represented by separate neural substrates that are specialized in processing affective information, as areas in right posterior occipital and superior temporal lobes exhibit face and voice selective patterns of activity that are not witnessed during ordinary face or voice detection (Kanwisher, McDermott, & Chun, 1997; Haxby, Hoffman, & Gobbini, 2000; Schirmer & Kotz, 2006). Additionally, these areas are critical to the initial recognition and identification of emotion, as listening to prosodic voices elicits activity in the middle superior temporal gyrus (mSTG) and viewing emotional faces has been associated with increased neural activity in the occipital face area (OFA) (Bradley, et al., 2003; Köchel, et al., 2011; Schirmer & Kotz, 2006). While these findings highlight the specialized processing that occurs for each modality, they do not examine the synergistic effect of these channels on affect perception.

Additionally, there has been considerable evidence to indicate that perceptual integration is associated with activity in the posterior STS, as this area is responds to both vocal and facial

sources of nonverbal affective information (Ethofer, Pourtois, & Wildgruber, 2006; Kreifelts, Ethofer, Shiozawa, Grodd, & Wildgruber, 2009; Watson, Latinus, Noguchi, Garrod, Crabbe, & Belin, 2014; Wright, Pelphrey, Allison, McKeown, & McCarthy, 2003). Overlapping affective face and voice recognition systems may indicate a fundamental neural mechanism, which is aimed at efficiently integrating affective and attentional information to facilitate recognition of and orienting towards potentially threatening environmental stimuli (Grandjean, et al., 2005). There has been considerable evidence to indicate that perceptual integration is associated with activity in the posterior STS, as this area is responds to both vocal and facial sources of nonverbal affective information (Ethofer, Pourtois, & Wildgruber, 2006; Kreifelts, Ethofer, Shiozawa, Grodd, & Wildgruber, 2009; Watson, Latinus, Noguchi, Garrod, Crabbe, & Belin, 2014). Overlapping affective face and voice recognition systems may indicate a fundamental neural mechanism, which is aimed at efficiently integrating affective and attentional information to facilitate recognition of and orienting towards potentially threatening environmental stimuli (Grandjean, et al., 2005).

Understanding the integration of these senses poses a particularly difficult problem as emotional information can be gleaned and identified in either modality independently, but their combined influence on affect perception is not clear. This perceptual quandary is best exemplified by the McGurk paradigm, which demonstrated that the fusion of two mismatched visual (spoken /ga-ga/) and auditory inputs (voiced /ba-ba/) could create an illusory percept (heard /da-da/) (McGurk & MacDonald, 1976). These findings been extended to affect perception research, by incorporating the use of affective face images and emotional vocalizations. These studies typically use an assortment of semantically neutral sentences voiced in different prosodies in conjunction with images from an array of morphed continua created from two oppositely valenced, static end-point face images (Massaro & Egan, 1996; de Gelder & Vroomen, 2000; Roberson, Damjanovic,

& Pilling, 2007). In this paradigm, the morphed continua enable the experimenter to assess the categorical perception of faces as it is presented with congruent and incongruent vocal stimuli. Faces on a morphed continuum are not perceived to change continuously by equal physical amounts, rather faces are seen as belonging to discrete emotional categories (Ekman & Friesen, 1976; Calder, Young, Perrett, Etcoff, & Rowland, 1996; Young, et al., 1997; de Gelder & Vroomen, 2000; Fujimura, et al., 2012). This perceptual schism is critical to studies of multimodal affect perception studies, wherein the perceptual boundary between two emotions can be tracked by the percentage of responses for each emotion at each place on the continua for each prosody (Fujimura, et al., 2012; Calder, Young, Perrett, Etcoff, & Rowland, 1996).

Several studies have indicated that emotional prosody produces verbal interference in face perception, as subjects' responses became biased towards the emotion expressed by prosody of the voice (de Gelder & Vroomen, 2000; Massaro & Egan, 1996; Campbell, 1996). Molholm and colleagues (2002) suggested that this verbal interference may indicate that the prosody of an affective vocal expression may modify the processing of emotional faces. Collectively, these findings suggest that perceptual acuity for facial and vocal displays of emotion is fundamental to human communication and that these two sensory inputs are independently represented by partially overlapping neuroanatomical areas.

This study will investigate the independent and joint effects which affective faces and voices have on emotion perception using functional near-infrared spectroscopy and two behavioral measures. Behavioral data will be acquired using a 2-alternative forced choice task to measure the percentage of responses that a stimulus was rated to be 'happy' or 'not happy' for

every face image in each prosody condition. The authors hypothesized that the prosodic voices would modify or produce a perceptual bias in participants' responses that could be evidenced by subjects' reaction times and two psychometric measures (the point of subjective equality (PSE) and the just noticeable difference (JND)). The percentage of happy responses for the happy prosody condition was expected to be greater than the other prosody conditions, with the angry prosody condition exhibiting the lowest percent happy responses, and the overall percent happy responses for the neutral condition falling between the two.

This study will attempt to dissociate the neural correlates of emotion perception by measuring changes in oxygenated-hemoglobin in response to affective faces, voices, and faces and voices paired together using fNIRS. This imaging modality was employed as fNIRS can provide additional insights into affective categorical perception paradigms by providing good spatial precision and temporal resolution (Gibson, Hebden, & Arridge, 2005). Additionally, while relatively few neuroimaging studies have focused on multimodal using nonverbal stimuli, an even smaller portion of this literature has been utilized fNIRS. Thus, implementation of fNIRS technique may provide additional insights into emotion perception in bimodal integration.

REFERENCES

Bartlett, M.S., Viola, P.A, Sejnowski, T.J., Golomb, B.A., Larsen, J., Hager, J.C., Ekman, P. (1996). Classifying facial action. *Advances in Neural Information Processing Systems.* D. Touretzky, M. Mozer, & M. Hasselmo (Eds.), MIT Press, pp. 823-829.

Belin, P., Bestelmeyer, P.E.G., Latinus, Watson, R. (2011). Understanding Voice Perception. *British Journal of Psychology* 102:711-725. doi:10.1111/j.2044-8295.2011.02041.x

Belin, P., Fillion-Bilodeau, S., Gosselin, F. (2008). The Montreal Affective Voices: a validated set of nonverbal affect bursts for research on auditory affective processing. *Behavioral Research Methods 40*(2):531-9. PMID: 18522064

Bradley, M.M., Sabatinelli, D., Lang, P.J., Fitzsimmons, J.R., King, W., Desai, P. (2003). Activation of the visual cortex in motivated attention. *Behav Neuroscience 117*(2):369-380. PMID: 12708533

Calder, A.J., Young, A.W., Perrett, D.I., Etcoff, N.L., Rowland, D. (1996). Categorical Perception of Morphed Facial Expressions. *Visual Cognition 3*(2): 81–118. doi:10.1080/713756735.

Campbell, R. (1996). Seeing Speech in Space and Time: Psychological and Neurological Findings. In *Spoken Language*, 3:1493–97. Philidelphia, United States: IEEE. doi:10.1109/ICSLP.1996.607899.

Darwin, C. (1872). *The expressions of emotion in man and animals*. London: John Murray.

de Gelder B., Vroomen, J. (2000). The Perception of Emotions by Ear and by Eye. *Cognition and Emotion 14*(3): 289–311. doi:10.1080/026999300378824.

Ekman, P., Friesen, W.V. (1976). *Pictures of Facial Affect*. Palo Alto, CA: Consulting Psychological Press.

Ekman, P. (1992). An Argument for Basic Emotions. *Cognition and Emotion* 3(4): 169–200. doi:10.1080/02699939208411068.

Ethofer, T., Pourtois, G., Wildgruber, D. (2006). Investigating audiovisual integration of emotional signals in the human brain. *Progress in Brain Research* 156:345-361. PMID: 17015090

Fitch, W.T. (2000). The evolution of speech: a comparative review. *Trends Cogn Science 4*(7):258-267. PMID: 10859570

Gibson, A.P., Hebden, J.C., Arridge, S.R. (2005). Recent advances in diffuse optical imaging. *Physics in Medicine and Biology 50*(4):R1-43. PMID: 15773619

Grandjean, D., Sander, D., Pourtois, G., Schwartz, S., Seghier, M.L., Scherer, K.R., Vuilleumier, P. (2005). The voices of wrath: brain responses to angry prosody in meaningless speech. *Nature Neuroscience 8*(2):145-156. PMID: 15665880

Haxby, J.V., Hoffman, E.A., Gobbini, M.I. (2000). The distributed human neural system for face perception. *Trends Cognitive Science 4*(6):223-233. PMID:10827445

Kanwisher, N., McDermott, J., Chun, M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J Neuroscience* 17: 4302±4311. PMID: 9151747

Köchel, A., Plichta, M.M., Schäfer, A., Leutgeb, V., Scharmüller, W., Fallgatter, A.J., Schienle, A. (2011). Affective perception and imagery: A NIRS study. *Int J Psychophysiology. 80*(3):192-197. PMID: 21419180

Kreifelts, B., Ethofer, T., Shiozawa, T., Grodd, W., Wildgruber, D. (2009). Cerebral representation of non-verbal emotional perception: fMRI reveals audiovisual integration area between voice- and face-sensitive regions in the superior temporal sulcus. *Neuropsychologia 47*(14):3059-3066. PMID: 19596021

Massaro, D.W., Egan, P.B. (1996). Perceiving affect from the voice and the face. *Psychonomic Bulletin Review 3*(2):215-21. doi: 10.3758/BF03212421. PMID: 24213870.

McGurk, H., MacDonald, J. (1976). Hearing lips and seeing voices. *Nature 264*(5588):746-8. PMID: 1012311.

Molholm, S., Ritter, W., Murray, M.M., Javitt, D.C., Schroeder, C.E., Foxe, J.J. (2002). Multisensory auditory-visual interactions during early sensory processing in humans: a high-density electrical mapping study. *Brain Res Cogn Brain Res*. *14*(1):115-28. PMID: 12063135.

Patel, S., Scherer, K.R., Björkner, Sundberg, J. (2011). Mapping emotions into acoustic space: The role of voice production. *Biological Psychology 87*(1):93-98. PMID: 21354259

Roberson, D., Damjanovic, L., Pilling, M. (2007). Categorical perception of facial expressions: Evidence for a "category adjustment" model. *Memory & Cognition 35*(7):1814-1829. PMID: 18062556

Scherer K.R., Johnstone T., Klasmeyer G. (2003). Vocal expression of emotion. In Davidson R.J., Scherer K.R., Goldsmith H.H. (Eds.), Handbook of affective sciences (pp. 433–456). New York: Oxford University Press.

Scherer, K.R., Kappas, A. (1988). Primate vocal expression of affective state. *Primate Vocal Communication* 171-194. doi: 10.1007/978-3-642-73769-5_13

Schirmer, A., Kotz, S.A. (2006). Beyond the right hemisphere: brain mechanisms mediating vocal emotional processing. *Trends in Cognitive Sciences 10*(1):24-30. PMID:16321562

Watson, R., Latinus, M., Noguchi, T., Garrod, O., Crabbe, F., Belin, P. (2014). Crossmodal Adaptation in Right Posterior Superior Temporal Sulcus during Face–Voice Emotional Integration. *J Neuroscience 34*(20):6813-6821. PMID: 24828635

Wright, T.M., Pelphrey, K.A., Allison, T., McKeown, M.J., McCarthy, G., (2003). Polysensory interactions along lateral temporal regions evoked by audiovisual speech. *Cerebral Cortex* 13: 1034–1043. PMID:12967920

Young, A. W., Rowland, D., Calder, A. J., Etcoff, N. L., Seth, A., Perrett, D. I. (1997). Facial expression megamix: Tests of dimensional and category accounts of emotion recognition. *Cognition*. 63:271-313. PMID: 9265872

# CHAPTER 2 – THE ROLE OF PROSODY IN BIMODAL AFFECT PERCEPTION

## 2.1 Introduction

*2.1.1 Affective Faces and Voices.* The ability to make inferences about the emotional state of another individual is facilitated by the decoding and identification of concurrently presented facial and vocal information. Deriving the contribution or role of each modality in affect perception is difficult as both faces and voices provide sufficient information to convey emotions independently (Ekman & Friesen, 1976; Schröder 2003; Belin, Fillion-Bilodeau, & Gosselin, 2008). Darwin (1872) was the first to recognize the innate correspondence of facial movements to the condition of one's internal emotional state. Ekman and Friesen (1976) expounded upon this concept, identifying six basic emotions (anger, happiness, sadness, disgust, fear, and surprise) which are associated with the stereotyped movements of groups of facial muscles (Bartlett, Viola, Sejnowski, Golomb, Larsen, Hager, & Ekman, 1996). These facial expressions appear to be highly salient, as performance on tests evaluating the recognition and characterization of others' emotional state is consistent across cultures ( Darwin, 1872; Ekman, 1993; Ekman, Friesen, & Wallace, 1971; Ekman & Frisen, 1976). This conservation implies that facial expressions are inherent and play an essential role in human communication. Yet, faces are rarely shown in silence, but are typically paired with a vocal counterpart.

The vocal system transmits emotional information through articulatory gestures which alter the rhythm, intensity, and intonation of a speaker's voice (Schröder, 2003; Grandjean, et. al, 2005). Prosody conveys emotion via the grouping of suprasegmental vocal features (fundamental frequency, rhythm, and amplitude) to produce distinct, affective vocalizations (Belin, Fillion-Bilodeau, & Gosselin, 2008; Juslin & Laukka, 2003; Patel, Scherer, Björkner, & Sundberg, 2011). These vocal qualities appear to be processed independently from linguistic cues as subjects can

accurately recognize and identify a speaker's mood when semantically neutral sentences are read in different prosodies (Johnson, Emde, Scherer, & Klinnert, 1986) or when spoken in a foreign language (Pell, Monetta, Paulmann, & Kotz, 2009; Scherer, Banse, & Wallbott, 2001; Thompson & Balkwill, 2006). Thus, prosody appears to be embedded in human communication and may be considered the vocal analog of the six basic emotions originally identified by Ekman & Friesen (Ekman & Friesen, 1976; Schröder, 2003). These findings make intuitive sense, as vocalizations are produced through the coordinated action of vocal and facial muscles which result in distinctive facial expressions (Schröder, 2003; Belin, Fillion-Bilodeau, & Gosselin, 2008). This congruency implies that vocalizations and facial expressions may be innately linked as visceral responses, communicating complementary affective information to the eyes and ears of an observer.

*2.1.2 Integration of Voices and Faces.* The instantaneous and automatic integration of visual information with vocal information is best illustrated during speech communication, where auditory information from a speaker's voice is directly linked to the movements of a speaker's face to form one coherent percept (McGurk & MacDonald, 1976). While the congruency of the messages is essential to interpersonal interactions, it does little to disentangle the efficacy of each modality in percept formation. One of the most influential experiments examining aberrant audiovisual integration is evidenced by the illusory percept formed (hearing /da-da/) by the fusion of two mismatched visual (spoken /ga-ga/) and auditory inputs (voiced /ba-ba/), commonly known as the McGurk Effect (McGurk & MacDonald, 1976). This phenomenon has been replicated in experiments of affect perception, where the vocal and facial cues of a stimulus contain conflicting emotional information (Massaro & Egan, 1996; de Gelder & Vroomen, 2000). These studies have shown that while both facial expressions and affective vocalizations can bias emotion perception (Massaro & Egan, 1996; de Gelder & Vroomen, 2000), faces appear to have the greatest effect

(Massaro & Egan, 1996; de Gelder & Vroomen, 2000; Abelin, 2007). However, the efficacy of each channel appears to vary as a function of the content of stimuli, instructions, and response directions (Massaro & Egan, 1996; de Gelder & Vroomen, 2000; Abelin, 2007). More generally, simultaneous presentation of conflicting visual and auditory input distorts perception that may hold special significance when integrating nonverbal emotional information.

Variations on the emotional McGurk experiment have provided keen insights into the integration of verbal and nonverbal affective vocal information with emotional faces. These studies typically employ a variation of semantically emotional sentences or affectively voiced prosodic stimuli paired with a morphed continuum created from two oppositely valenced, static end-point face images (Massaro & Egan, 1996; de Gelder & Vroomen, 2000; Roberson, Damjanovic, & Pilling, 2007). Facial expressions are perceived categorically (Ekman & Friesen, 1976; Calder, Young, Perrett, Etcoff, & Rowland, 1996; Young, et al., 1997; de Gelder & Vroomen, 2000; Fujimura, et al., 2012), which enables researchers to quantify the changes in perception that occur as stimuli change in equal physical amounts across a continuum. Interestingly, humans do not perceive morphed continua continuously, but perceive stimuli as belonging to one of two discrete categories, they exhibit categorical perception (Harnad, 1987). This phenomenon is pertinent to studies of audiovisual integration, where the perceptual boundary between two emotions can be tracked by a subject's identification responses, (Fujimura, et al., 2012; Calder, Young, Perrett, Etcoff, & Rowland, 1996) with faces nearest to the center of the morph continuum being the hardest to identify (Calder et al., 1996).

The perceptual boundary between emotional categories can be quantified using two psychophysical measures, the point of subjective equality (PSE), associated with identification and the just noticeable difference (JND), related to discrimination. For the purposes of this study, the

PSE will be defined as the point at which a stimulus is equally likely to be judged as happy or not happy. The JND is a percentage value of the amount of physical change needed to discriminate between two stimuli 50% of the time. The magnitude of the JND indicates the variance in subject responses, which can be equated to the level of confusion participants experience in each condition.

These measures will be used to assess the cross-modal effects of simultaneously presented affective vocal and facial expression in affect perception, with the intent of adding to a literature of experiments using nonverbal stimuli. Regardless, several studies have suggested that emotional prosody interferes with face perception, as subjects' identification of facial expressions becomes biased towards the emotion expressed in the vocal utterance (de Gelder & Vroomen, 2000; Massaro & Egan, 1996; Pourtois, de Gelder, Vroomen, Rossion, & Crommelinck, 2000; Campbell, 1996) and this effect persists even when instructed to ignore the auditory stimuli (de Gelder & Vroomen, 2000). Further, verbal interference appears to degrade the categorical perception of faces more than interference with incongruent faces during a vocal categorical perception task (Roberson & Davidoff, 2000). Findings from de Gelder & Vroomen (2000) indicated that the perception of affective faces is biased in the direction of a simultaneously presented prosodic voice, and that the impact of the voice increases as the emotions of the facial expressions become more ambiguous (Vroomen et al., 2001). Molholm and colleagues (2002) suggested that these data indicate that early processing of visual inputs is modified by auditory inputs. Massaro & Egan (1996) found that while both facial expressions and affective vocal cues are effective in biasing responses from happy to angry, faces appear to exert a greater influence in bimodal integration emotion perception. Similarly, other findings have suggested that faces appear to play a greater role in biasing affect perception in bimodal conditions (Hess, Kappas, & Scherer, 1988), however this may vary by age (Bugenthal, Kaswan, Love, & Fox, 1970), emotion (Li, et al., 2013),

directions (de Gelder & Vroomen, 2000), choice of stimuli, and subject characteristics (Massaro & Egan, 1996). Nevertheless, multiple studies have shown that bimodal integration of coherent stimuli produces gains in response accuracy and decreases reaction times when identifying emotions (de Gelder & Vroomen, 2000; Pell, 2005; Ethofer, Bretscher, Gschwind, Kreifelts, Wildgruber, & Vuilleumier, 2012).

*2.1.3 The Current Study*. This study will investigate emotion perception as it is conveyed via faces, voices, and faces and voices paired together. The authors have elected to use short nonverbal affective bursts as vocal stimuli, as they are paralinguistic to minimize any cognitive processing that may indirectly bias participants' attention to the face or voice (Schröder, 2003; Belin, Fillion-Bilodeau, & Gosselin, 2008). Additionally, these stimuli may be more ecologically valid as they were evoked by intense emotions which may have caused actors' articulatory patterns and facial gestures to be more like that of natural emotions (Schröder, 2003; Belin, Fillion-Bilodeau, & Gosselin, 2008). This study will employ a two-alternative forced choice task where subjects will be instructed to indicate if the emotion they perceived for each trial is "happy" or "not happy" with no reference to attend to the voice or face. This was done to limit any inherent bias to respond to one modality or the other.

We predict that reaction times will be slowest for all conditions when they are at the category boundary or most ambiguous portion of the continuum (Massaro & Etcoff, 1996; de Gelder & Vroomen, 2000). Further, we hypothesize that reaction times for each prosody condition will vary as a function of their congruency with the emotion of the simultaneously presented face, with reaction times being faster when the faces and voices express the same emotion and slower when they are mismatched (de Gelder & Vroomen, 2000). The authors predict that the PSEs for each condition will be biased in the direction of the simultaneously presented prosody, as an

indication in a perceptual shift in the identification curve. JND values will be interpreted as the level of confusion in subjects' responses, which together with the reaction time data can be used as indicators of how well defined the perceptual boundaries are between the two emotion categories.

## 2.2 Methods

*2.2.1 Participants.* Thirty undergraduate students (15 female) from Colorado State University participated in this study. The mean age for participants was 21.03 (3.35) years. All participants filled out three brief questionnaires regarding general and mental health, as well as drug and alcohol use. The Duke health profile (DUKE) was used to gauge subjects' perceived level of physical, mental, social, and overall health (Parkerson, Broadhead, & Tse, 1990). One subject chose not to complete the DUKE general health questionnaire; this is indicated under the questionnaires completed column. Alcohol use was measured using the Alcohol Use Disorders Identification Test (AUDIT; Babor, de la Fuente, Saunders, & Grant, 1992), and drug use was assessed using the Drug Abuse Screening Test (DAST-10; Skinner, 1982). Questionnaire results and scoring cutoffs are shown in table 2.1.

**Table 2.1** *Mean Questionnaire Scores Concerning Drug and Alcohol Use, and General Physical and mental health.*

| Scale | Variables | | Age | | Score | |
|---|---|---|---|---|---|---|
| | | | *M* | *SD* | *M* | *SD* |
| DUKE | | *n* = 29(15) | 21.10 | 3.38 | | |
| *Health Measures* | | | | | | |
| | Physical health | | | | 83.45 | 13.17 |
| | Mental health | | | | 82.50 | 15.78 |
| | Social health | | | | 84.83 | 15.26 |
| | General health | | | | 83.57 | 9.99 |
| | Perceived health | | | | 83.93 | 23.78 |
| | Self-esteem | | | | 87.50 | 11.10 |
| *Dysfunction Measures* | | | | | | |
| | Anxiety | | | | 20.69 | 14.88 |
| | Depression | | | | 22.50 | 17.56 |
| | Anxiety-depression | | | | 19.39 | 14.53 |
| | Pain | | | | 17.24 | 24.19 |
| | Disability | | | | 3.45 | 12.89 |
| DAST-10 | | *n* = 30(15) | 21.03 | 3.35 | | |
| | Drug abuse | | | | 1.74 | 1.44 |
| AUDIT | | *n* = 30(15) | 21.03 | 3.35 | | |
| | Alcohol use | | | | 3.84 | 2.85 |

Note: Parentheses indicate number of female participants. Scores for the DUKE are raw scores from a scale of 0.0-100.0. High scores for health measures indicate good health, high scores for the dysfunction measures equates to poor health. DAST-10 contains 10 items with scores ranging from 0.0-10.0, lower scores (1-5) indicating lower to moderate drug use, and higher scores (6-10) suggesting substantial to severe drug use. Total AUDIT scores greater than 8 indicate dangerous and harmful alcohol consumption, with scores ranging from 0.0-40.0.

*2.2.2 Auditory Stimuli*. Auditory stimuli were taken from the freely available Montreal Affective Voices database (Belin, Fillion-Bilodeau, & Gosselin, 2008), in which professional actors produced short, nonverbal affective interjections of the vowel /a/, which sounds similar to the a in “ah.” Three vocalizations expressed in angry, happy, and neutral prosody were chosen for each actor (two actors (one male)) resulting in a total of six vocalizations, previously matched and

validated for valence (negative, positive), arousal, and perceived intensity (Belin, Fillion-Bilodeau, & Gosselin, 2008).

*2.2.3 Facial Stimuli*. Face stimuli were taken from the NimStim database (Tottenham, Tanaka, Leon, McCarry, Nurse, Hare, Marcus, Westerlund, Casey, & Nelson, 2009). One angry and one happy closed-mouth image were selected from a subset of 20 actors (10 men). Images were grayscaled and cropped tightly around the face so that no hair, neck or clothing was visible. These features were cropped as they are often greatly distorted during the morphing process which may have pulled focus away from the model's facial expression (Tottenham, et al., 2009). One continua were generated for each actor, using Psychomorph software (Tiddeman, Burt, & Perrett, 2001; Tiddeman & Perrett, 2002). Each continuum consisted of two end-point prototype images (angry or happy), which were morphed together in seven steps (two endpoints and 5 morphs, in 12.5% steps) so that the mid-point image would be a 50% combination of each prototype image.

*2.2.4 Experiment.* Auditory and facial stimuli were presented alone or paired together in a bimodal (face and prosody) condition to create seven different stimulus conditions (Figure 1.1, top).

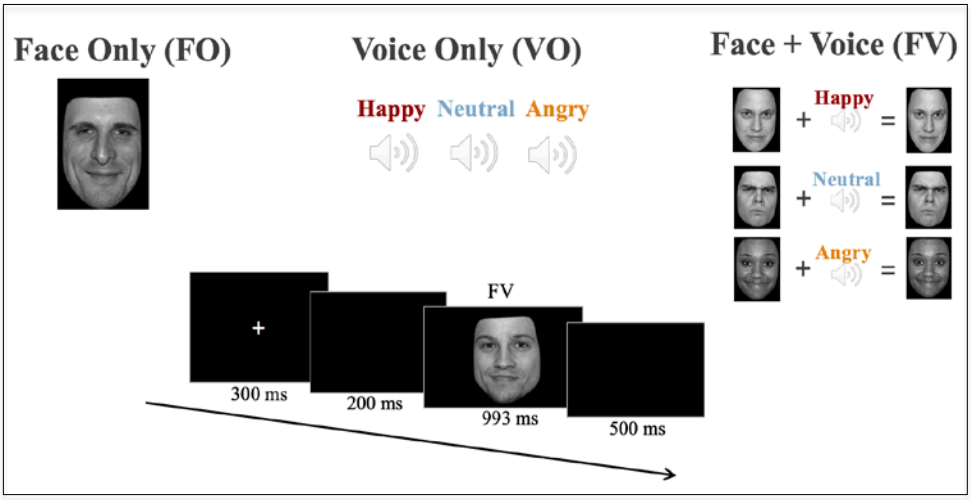


**Figure 2.1:** (Top) Examples of face, voice, and face and voice stimuli. (Bottom) Example of a single trial with a face + voice (FV) stimulus.

Voice only conditions consisted of an affective voice presented with a white fixation cross on a black screen. Participants received binaural auditory stimulation (70 dB SPL) through EAR 3a foam insert earphones. Morphed face stimuli were either shown in silence or simultaneously presented with an auditory stimulus on a monitor located 45 cm in front of the subject. Face stimuli subtended 7.62 degrees of vertical visual angle and 5.72 degrees of horizontal angle. Each trial began with a white fixation cross on a black background which lasted 300 ms, proceeded by a 200 ms pause, after which, a voice, face, or voice and face were presented for 994 ms. A black screen appeared for 500 ms after the stimulus had ended, creating trials which totaled 1994 ms in duration (Figure 2.1, bottom). Subjects were instructed to identify the emotion expressed by the actor for every trial in a 2-alternative (happy/not happy) forced choice procedure using the left and right bumpers of an Xbox controller with no specific reference to the face or voice. Subjects were asked to respond quickly, to ensure that their responses would occur within the stimulus window.

Stimuli were presented in a hybrid block design presented in E-Prime 1.0. There was a total of 420 trials (20 actors x 7 conditions x 7 faces on a continuum) with trials grouped into pseudo-random voice condition blocks (Figure 2.2). All blocks consisted of 14 trials. Block conditions

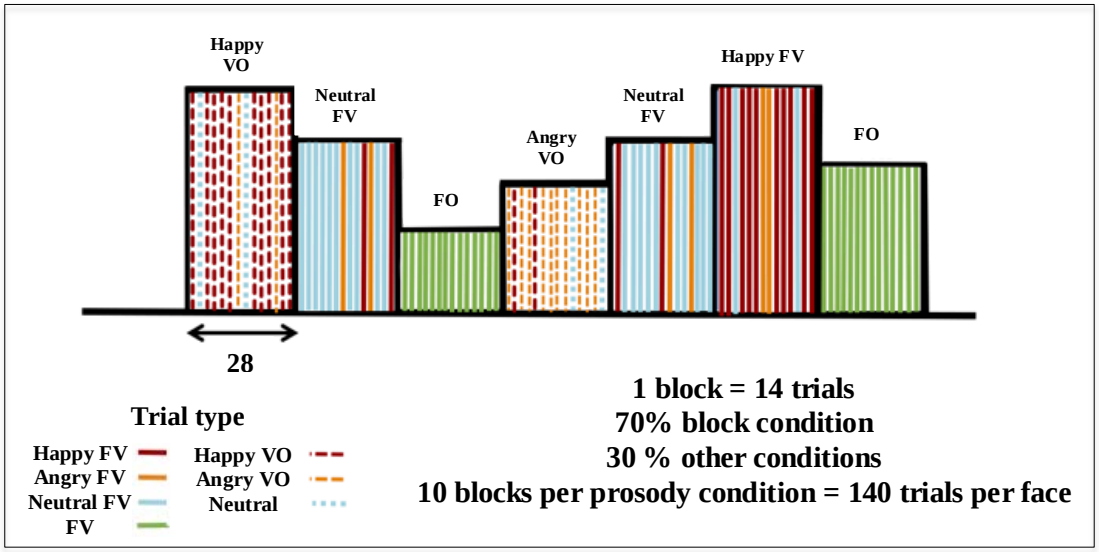


**Figure 2.2:** Schematic of trial and block organization.

were named after the trial type which made up 70% of the trials. The remaining 30% of trials was divided equally between the two remaining voice conditions. Face only blocks contained 14 randomly selected morphed faces. Each condition (voice only (VO), face only (FO), or face+voice (FV)) had 10 blocks. All 70 blocks were 28s long for a total duration of 32.7 minutes. Faces and vocal stimuli were matched by gender. While the experimental paradigm included seven conditions only four (FO, happy FV, angry FV, neutral FV) were used for data analysis. Participants were informed that they would be presented with affective faces and voices which would be shown alone (FO, VO) or paired together (FV), and that for each trial, regardless of stimulus type they were to indicate if the emotion they perceived was "happy" or "not happy" by pressing the left or right bumper on an Xbox controller. Button press responses were analyzed for the proportion of happy responses and reaction times for each prosody condition and face. Reaction times were measured at the onset of each stimulus presentation. Reaction times were excluded if they were less than 200 ms or greater than 994 ms.

### 2.3 Results

Percent happy responses were analyzed using a two-way within subjects ANOVA, with condition (angry FV, neutral FV, happy FV, face only) and face step (faces 1-7 on continuum) as within-subjects variables. The data exhibited significant main effects for condition $F(3, 84) = 13.58$, $p = .000$, $\eta_p^2 = .327$, face step $F(6, 168) = 364.30$, $p = .000$, $\eta_p^2 = .929$, as well as, the interaction between condition and face step $F(18,504) = 9.78$, $p = .000$, $\eta_p^2 = .259$.1 All post hoc comparisons were made using Bonferroni adjusted Fisher's LSD. The happy FV condition exhibited the highest proportion of happy responses (.521 ± .019) compared to the FV angry (.373 ± .020, $p = .000$), neutral FV (.426 ± .021, $p = .003$), and face only (.424 ± .012, $p = .000$) conditions. Proportion of happy responses for the angry FV condition (.373 ± .020), were

significantly lower than both the neutral FV condition (.426 ± .021, $p$ = .006) and face only (.424 ± .012, $p$ = .021) condition. The proportion of happy responses increased for all conditions across the face continuum (Figure 2.3). To analyze the hypothesized bias effects, data were fit using a logistic function to calculate point of subjective equality (PSE) and just noticeable difference

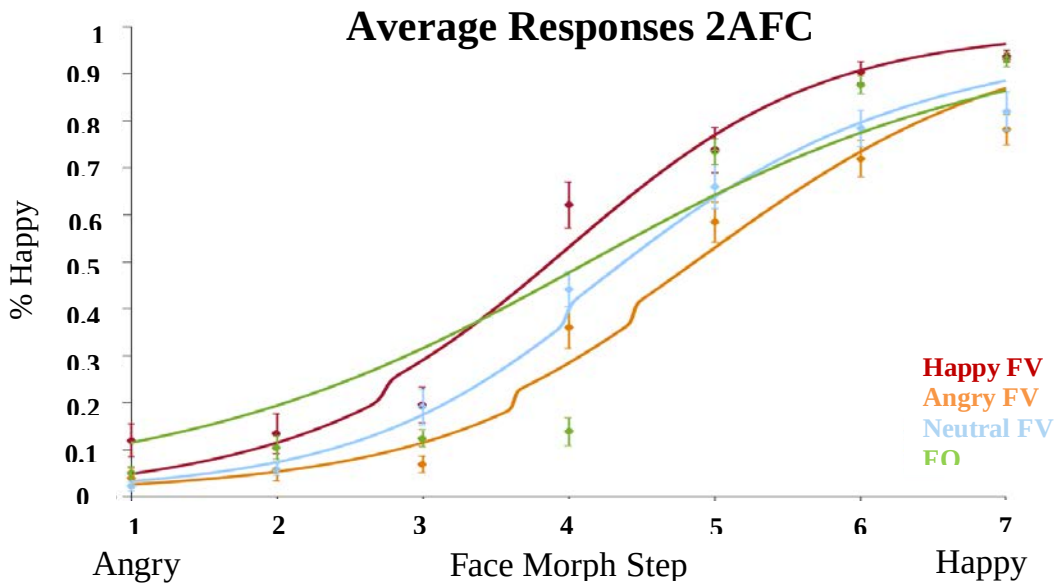


**Figure 2.3:** Proportion of happy responses for each step in the face morph continuum when combined with each condition.


**Figure 2.4:** Mean reaction times across each face step for each condition.

(JND) values, which were analyzed using two identical one-way repeated measures ANOVAs with condition (FV prosody, face) as the within subjects factor. Results showed that there was a statistically significant difference in PSE values between conditions $F(3,115) = 8.48$, $p = .000$ with the happy FV condition having a significantly different PSE (3.78 ± .946) than both the angry (4.87 ± 1.14, $p = .000$) and neutral (4.70 ± .966, $p = .000$) FV conditions. The PSE for the face only condition (4.15 ± .628) was also significantly different than the neutral ($p = .028$) and angry ($p = .003$) FV conditions, but it was not significantly different from the happy FV condition ($p = .132$) (Figure 2.5).

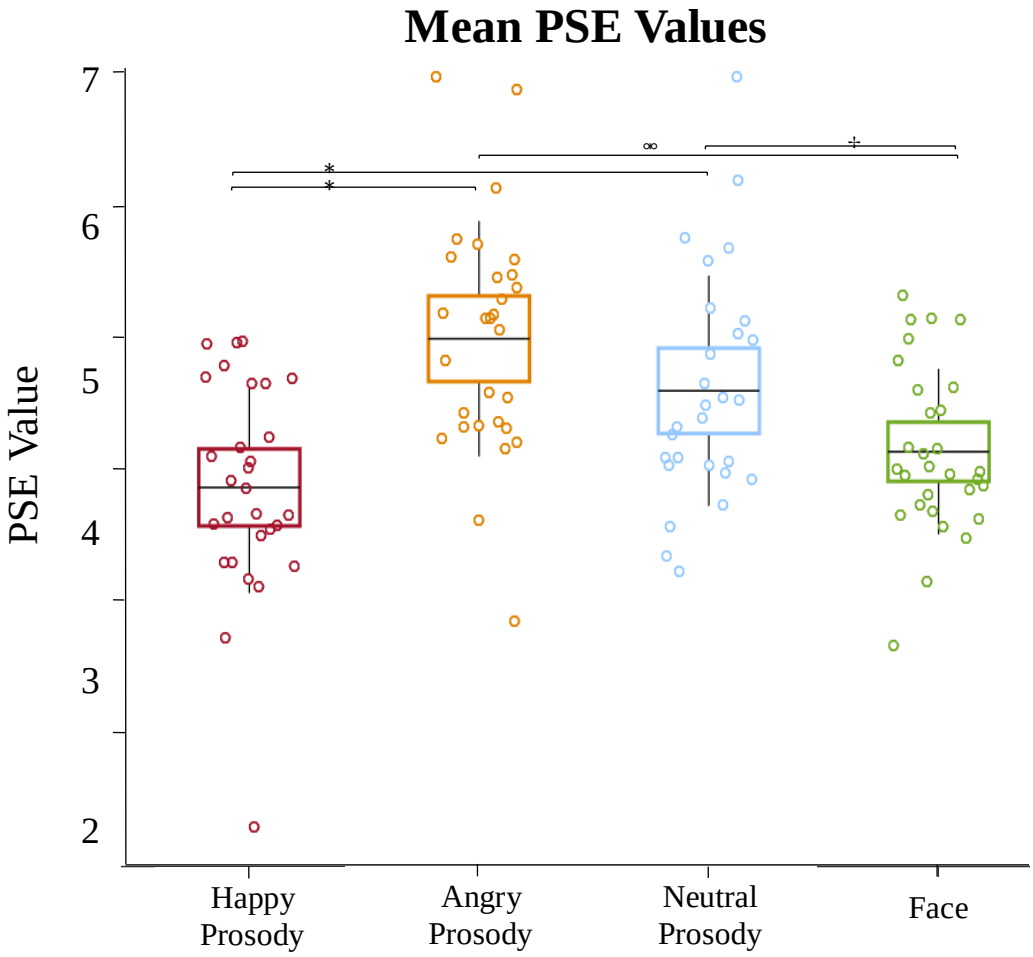


**Figure 2.5:** Group means and individual point of subjective equality (PSE) values. Lower PSE values indicate that the stimuli were perceived to be happier than higher PSE values. Significance values indicated by p <.000 = *, p < .005 = □, p < .05= ✢

The ANOVA for JND revealed a statistically significant difference in JND values between conditions $F(3,115) = 13.39$, $p = .000$. The face only condition exhibited a significantly larger JND (3.32 ± .939) than all FV conditions (happy, 1.77 ± .689, $p = .000$; angry, 2.04 ± 1.20, $p = .000$; neutral, 2.16 ± 1.20, $p = .000$). The JND for the happy FV condition was not significantly different

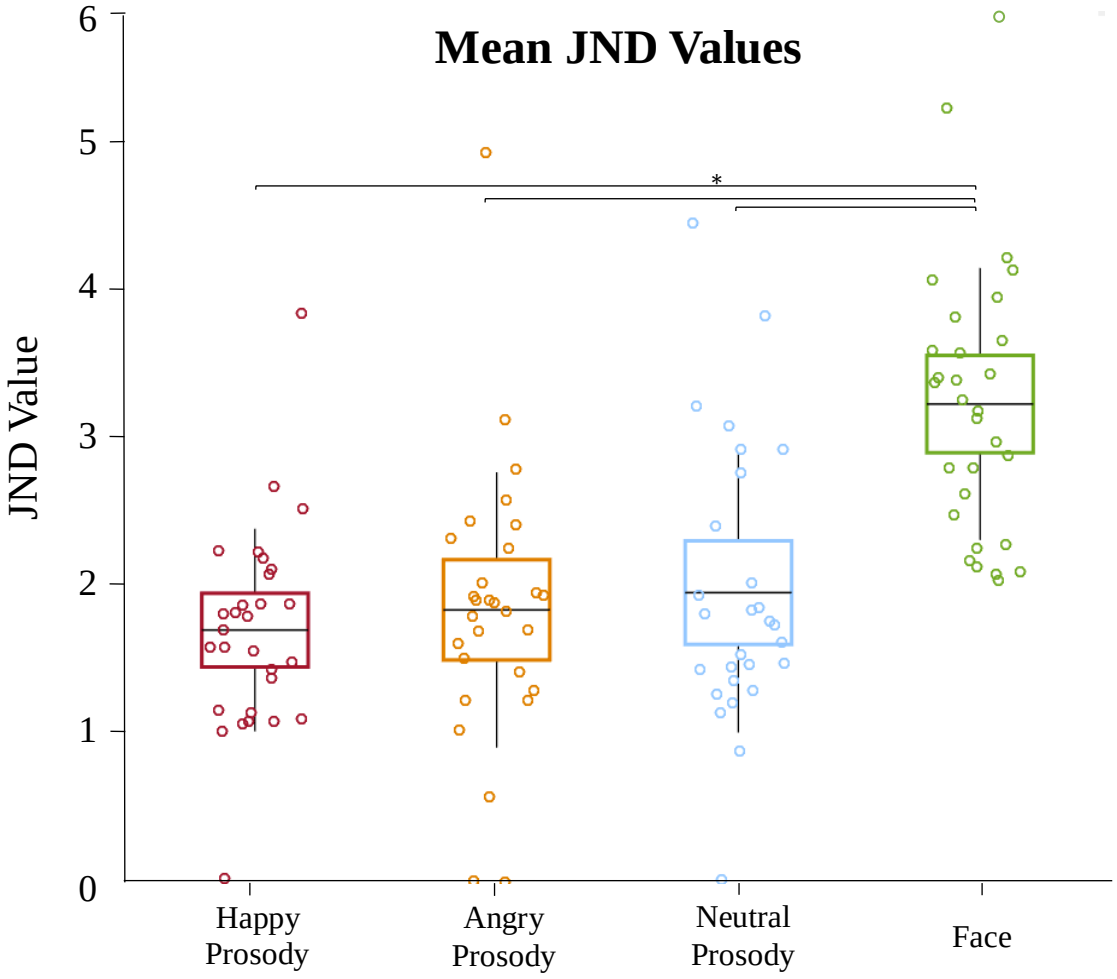


**Figure 2.6:** Group means and individual just noticeable difference (JND) values. Lower values indicate lower variability in subjects' responses. Significance values indicated by p <.000 = *.

from the angry ($p = .313$) or neutral ($p = .149$) FV conditions, which also did not significantly differ from one another ($p = .656$), see figure 2.6.

Subject reaction times were analyzed using an ANOVA with the same within subjects factors and levels, which revealed a significant main effect for face step $F(6,162) = 2.89$, $p = .010$,

$\eta_p^2$ = .301, condition $F(3,81) = 20.01$, $p = .000$, $\eta_p^2$ = .426, and a significant interaction between face step and condition $F(18,486) = 3.15$, $p = .000$, $\eta_p^2$ = .105. Pairwise comparisons revealed significantly faster reaction times for the face only condition when compared to all FV conditions (happy, 709.53 ± -51.08, $p = .000$; angry, 709.53 ± -40.57, $p = .000$; neutral, 709.53 ± -44.20, $p = .000$). There were no significant differences in reaction times between the FV conditions.

While there were no significant differences in the average reaction times for each FV condition, three one-way within subjects ANOVAs, one for each FV condition, with face step as the within subjects factor were used in an exploratory analysis to see if gains in reaction time across the continuum varied as a function of the emotional congruency of the face and voice. Reaction times for the happy FV condition ($F(6,199) = .774$, $p = .591$) and the neutral FV condition ($F(6,196) = 1.51$, $p = .177$) were not significantly different across the face continuum. However, while the ANOVA for the angry FV condition did not reach significance ($F(6,203) = 2.02$, $p = .064$), post hoc comparisons revealed that reaction times for the first face step were significantly faster than steps five (731.39 ± -46.25, $p = .05$) and six (731.39 ± -46.35, $p = .05$) and reaction times for the second step were significantly faster than steps four (718.28 ± -55.87, $p < .05$), five (718.28 ± -59.36, $p < .01$), and six (718.28 ± -59.46, $p < .01$).

### 2.4 Discussion

Emotion perception is the seemingly automatic integration of vocal and facial cues that together form a whole percept. While affective information can be gleaned from and identified in either modality independently, the relative contribution and interaction of these channels is unclear. This study implemented a 2-alternative forced choice task to examine the effect of unimodal (face only) and bimodal (face and voice (FV)) stimuli in biasing affect perception. Subjects were presented with affective bursts voiced in three prosodies (angry, happy, and neutral)

and images of faces morphed on a 7-step continuum created from two end-point images, one angry and one happy. These stimuli were combined to form seven conditions in which subjects were presented with either a single face image, a voice, or a face paired with a prosodic voice. Of these seven conditions, four were used to calculate subjects' PSE and JND values, and reaction times. Button-press responses were used to indicate if the emotion they perceived was "happy" or "not happy" on every trial, with no reference to the face or voice.

All four (happy FV, angry FV, neutral FV, face only) conditions exhibited a significant increase in the proportion of 'happy' responses as a function of the location of each face on the morphed continuum (Figure 2.3). This pattern may reflect the twofold saliency of happy faces in affect perception (Massaro & Egan, 1996; de Gelder et al., 1998) in that subjects' responses not only paralleled the increase in happy facial features across the face continuum, but this occurred across stimulus conditions, in the absence of any directions to pay attention to the face. Interesting, a similar pattern of responses was produced during a bimodal task when subjects were explicitly told to pay attention to an actor's face (de Gelder & Vroomen, 2000). The happy FV condition exhibited a significantly higher proportion of happy responses than all other conditions. The angry FV condition had the lowest proportion of happy responses, with the neutral FV and face only conditions having significantly higher percentages of happy responses than the angry condition, but significantly lower scores than the happy condition. The neutral FV and face only conditions were not significantly different from one another.

The face only condition had significantly faster reaction times than all FV conditions, which exhibited the slowest reaction times at the midpoint of the face continuum (Figure 2.3). This may suggest that concurrently presented facial and vocal cues require more processing time than faces alone to resolve ambiguity (Massaro & Egan, 1996; Pell, 2005). The neutral FV condition

exhibited an inverted 'U' shape, with fastest reaction times occurring at both end-points. While there were no significant differences in overall reaction times between FV conditions, an exploratory analysis was used to examine the effect of emotional congruency between faces and voices on reaction times for each prosody condition (de Gelder & Vroomen, 2000; Pell, 2005). Interestingly, only the angry FV condition exhibited significant increases in reaction times with a drastic change occurring between the second and fourth step, a pattern resembling the categorical boundary between the angry and happy faces in the proportion happy responses data (Figure 2.3). The happy FV condition did not display a similar reversed pattern of responses, as would be expected from equally arousing, but oppositely valenced stimuli (Calder, et al., 1996; Etcoff & McGee, 1996; Feldman, Barrett, & Russell, 1998; de Gelder & Vroomen, 2000). Together, this may indicate that angry FV combination may have been processed differently than the happy and neutral FV, with faster processing for emotionally congruent angry faces. These gains and losses in reaction time might be related to a "negativity bias", an innate tendency to attend, respond, and identify negatively valenced stimuli over other emotions (Horstmann & Bauland, 1996; Öhman, Lundqvist, & Esteves, 2001; Dijksterhuis & Aarts, 2003; Nasrallah, Carmel, & Lavie, 2009).

The happy FV and face only conditions had PSEs that occurred significantly earlier in the continuum than the neutral and angry FV conditions, indicating that happy prosodic voices appear to shift subjective judgments of affect perception so that morphed faces will appear to be 'happier' than their physical composition. Additionally, there were no significant differences between the happy FV and face only condition, nor the neutral and angry FV conditions (Figure 2.5). The face only and happy conditions may have had very similar PSEs due to the highly salient nature of happy faces, which may not have benefited from the presence of a prosodic voice. Additionally, these null results may be partially related to the editing performed on the neutral prosodic stimuli,

which was required for the original design of this experiment to be used in a neuroimaging study. Such trimming may have affected the ability of each sound clip to convey emotion (Belin, Fillion-Bilodeau, & Gosselin, 2008), as prosody recognition rates vary as a function of emotion and duration (Pell, 2005; Cornew, Carver, & Love, 2009). Alternatively, these data could suggest an inherent negativity bias which made the voices for the neutral and angry FV conditions appear angrier than the happy FV and face only conditions. Further, a negativity bias may account for a potential ceiling effect that prevented the angry FV condition from being perceived as 'angrier' than the neutral FV condition

JND values were smallest for the happy prosody condition, which may be related to the saliency of happy faces when paired with prosodic voices (Figure 2.6). However, this effect did not carry over to the face only condition, which had the largest JND, indicating that faces shown alone were the most confusing, producing the greatest variation in subjects' responses. These findings are somewhat surprising as the face only condition had reaction times that were significantly faster than all prosody conditions. This may suggest that while processing speed may be faster for unimodal stimuli, face and voice appears to increase sensitivity to differences in emotional expression, at least for happy prosody. The magnitude of the JND for faces also calls into question the subjects' ability to recognize the both the angry and happy physical traits carried by the face morphs. If these two emotions were equally arousing and oppositely valenced it would seem as though the category boundary would be as well-defined for happy and angry, but the categorical perceptual boundary appears more strongly to happy prosody in our data, suggesting that the angry faces may have been more emotionally ambiguous to our participants.

While a negativity bias for paying attention to angry stimuli has been reported in the literature (Horstmann & Bauland, 1996; Öhman, Lundqvist, & Esteves, 2001; Dijksterhuis &

Aarts, 2003), the applicability of findings from the current study are unclear. Given this discordance, it is important to evaluate both the vocal and facial stimuli used in this study, as it is imperative that both emotions be equal in arousal and intensity, and oppositely valenced. However, this dynamic can be inherently difficult when using positively and negatively valenced stimuli (Tottenham, et al., 2009), especially under conditions where the emotional properties and overall intensity of each emotion ('hot anger' verses 'cool anger') are poorly defined or differentially produced (posed versus evoked), increasing variation in actors' portrayals (Gur, et al., 2002; Schröder, 2003). The use of closed mouth stimuli in the current study have limited the evocative power of the chosen emotions as facial expressions become slightly less arousing and recognizable when mouths are closed, and this effect appears to more detrimental to angry than happy faces (Tottenham, et al., 2009). Thus, a limitation of the current study may have been that the angry faces were not perceived as 'angry enough', which may account for the differences in reaction times between the angry and happy prosody conditions, and the larger JND in the face only condition.

Overall, the findings of this study fit within a broader literature of affect perception research (Massaro & Egan, 1996; Etcoff & McGee, 1996; de Gelder & Vroomen, 2000; Pourtois, et al., 2000; Molholm et al. 2002; Campbell, 1996) which has demonstrated that auditory inputs modify the processing of visual stimuli. The present study has shown that prosody is highly salient and that the processing of bimodal stimuli can be biased in the direction of a simultaneously presented affective voice, at least when the prosodic channel is happy. Additionally, this study adds to the current literature in showing that while two emotions may be oppositely valenced, and exhibit similar intensity ratings, the full efficacy of these emotions and sensory channels may

depend on the stimuli or directions of the experiment (Massaro & Egan, 1996, de Gelder & Vroomen, 2000).

REFERENCES

Babor, T.F., de la Fuente, J.R., Saunders, J., Grant, M. (1992). AUDIT. *The Alcohol Use Disorders Identification Test. Guidelines for use in primary health care*. Geneva, Switzerland: World Health Organization.

Banse R, Scherer KR. (1996). Acoustic Profiles in Vocal Emotion Expression. *J Personality & Social Psychology 70*(3): 614-636. PMID: 8851745

Bartlett, M.S., Viola, P.A, Sejnowski, T.J., Golomb, B.A., Larsen, J., Hager, J.C., Ekman, P. (1996). Classifying facial action. *Advances in Neural Information Processing Systems.* D. Touretzky, M. Mozer, & M. Hasselmo (Eds.), MIT Press, pp. 823-829.

Belin, P., Fillion-Bilodeau, S., Gosselin F. (2008). The Montreal Affective Voices: a validated set of nonverbal affect bursts for research on auditory affective processing. *Behavioral Research Methods 40*(2):531-9. PMID: 18522064

Bugenthal, D.E., Kaswan, J.W., Love, L.R., Fox, M.N. (1970). Child versus adult perception of evaluative messages in verbal, vocal, and visual channels. *Developmental Psychology* 2:367-375. doi.org/10.1037/h0029166

Calder, A.J., Young, A.W., Perrett, D.I., Etcoff, N.L., Rowland, D. (1996). Categorical Perception of Morphed Facial Expressions. *Visual Cognition 3*(2): 81–118. doi:10.1080/713756735.

Campbell, R. (1996). Seeing Speech in Space and Time: Psychological and Neurological Findings. In *Spoken Language* 3:1493–97. Philidelphia, United States: IEEE. doi:10.1109/ICSLP.1996.607899.

Cornew, L., Carver, L., Love, T. (2009). There's more to emotion than meets the eye: A processing bias for neutral content in the domain of emotional prosody. *Cognition and Emotion 24*(7):1133-1152. PMID: 21552425.

Darwin, C. (1872). *The expressions of emotion in man and animals.* London: John Murray.

de Gelder, B., Vroomen, J. (2000). The Perception of Emotions by Ear and by Eye. *Cognition and Emotion 14*(3): 289–311. doi:10.1080/026999300378824.

Dijksterhuis, A., Aarts, H. (2003). On wildebeests and humans: The preferential detection of negative stimuli. *Psychological Science* 14:14–18. PMID: 12564748

Ekman, P., Friesen., Wallace, V. (1971). Constants across cultures in face and emotion. *Journal of Personality and Social Psychology 17*(2),124-129. PMID: 5542557

Ekman, P., Friesen, W.V. (1976). *Pictures of Facial Affect.* Palo Alto, CA: Consulting Psychological Press.

Ekman, P. (1992). An Argument for Basic Emotions. *Cognition and Emotion 3*(4): 169–200. doi:10.1080/02699939208411068.

Ekman, P. (1993). Facial expression and emotion. *American Psychologist.* 48, 384-392.

Etcoff, N. L., Magee, J. (1992). Categorical perception of facial expressions. *Cognition* 44**,** 227-240. PMID: 1424493

Ethofer, T., Bretscher, J., Gschwind, M., Kreifelts, B., Wildgruber, D., Vuilleumier, P. (2012). Emotional voice areas: anatomic location, functional properties, and structural connections revealed by combined fMRI/DTI. *Cerebral Cortex 22*(1):191-200. PMID: 21625012.

Fujimura, T., Matsuda, Y.T., Katahira, K., Okada, M., Okanoya, K. (2012). Categorical and dimensional perceptions in decoding emotional facial expressions. *Cognition and Emotion 26*(4):587-601. PMID: 21824015

Gur, R.C., Sara, R., Hagendoorn, M., Marom, O., Hughett, P., Macy, L., Turner, T., Bajcsy, R., Posner, A., Gur, R.E. (2002). A method for obtaining 3-dimensional facial expressions and its standardization for use in neurocognitive studies. *J Neuroscience Methods 115*(2):137-43. PMID: 11992665.

Hess, U., Kappas, A., Scherer, K. (1988). Multichannel communication of emotion: synthetic signal production. In Scherer, K. (Ed), Facets of emotion: Recent research, 161-182.

Horstmann, G., Bauland, A. (1996). Search asymmetries with real faces: Testing the anger-superiority effect. *Emotion* 6:193–207. PMID: 16768552

Johnson, W.F., Emde, R.N., Scherer, K.R, Klinnert, M.D. (1986). Recognition of emotion from vocal cues. *Archives of General Psychiatry* 43, 280-283. PMID: 3954549

Laukka, P. (2005). Categorical perception of vocal emotion expressions. *Emotion 5*(3):277-95. PMID: 16187864.

Li, A., Fang, Q., Jia, Y., Duang, J. *Chinese Computational Linguistics and Natural Language Processing Based on Naturally Annotated Big Data*: *Emotional McGurk Effect? A Cross-Cultural Investigation on Emotion Expression under Vocal and Facial Conflict*., edited by M. Sun et al.: Springer-Verlag Berlin Heidelberg, 2013, pp. 214–226, DOI: 10.1007/978-3-642-41491-6_20.

Massaro, D.W., Egan, P.B. (1996). Perceiving affect from the voice and the face. *Psychonomic Bulletin Review 3*(2):215-21. doi: 10.3758/BF03212421. PMID: 24213870.

McGurk, H., MacDonald, J. (1976). Hearing lips and seeing voices. *Nature 264*(5588):746-8. PMID: 1012311.

Molholm, S., Ritter, W., Murray, M.M., Javitt, D.C., Schroeder, C.E., Foxe, J.J. (2002). Multisensory auditory-visual interactions during early sensory processing in humans: a high-density electrical mapping study. *Brain Res Cogn Brain Res*. *14*(1):115-28. PMID: 12063135.

Nasrallah, M., Carmel, D., Lavie, N. (2009). Murder, She Wrote: Enhanced Sensitivity to Negative Word Valence. *Emotion 9*(5), 609–618. PMID: 19803583

Öhman, A., Lundqvist, D., Esteves, F. (2001). The face in the crowd revisited: A threat advantage with schematic stimuli. *J Personality and Social Psychology* 80:381–396. PMID: 11300573

Parkerson, G.R., Broadhead, W.E., Tse, C.K. (1990). The Duke Health Profile: A 17-item measure of health and dysfunction. *Med Care. 28*(11), 1056-1072. PMID: 2250492.

Pell, M.D. (2005). Nonverbal Emotion Priming: Evidence from the Facial Affect Decision Task. 29(1): 45–73. doi:10.1007/s10919-004-0889-8.

Pell, M.D., Monetta, L., Paulmann, S., Kotz, S.A. (2009). Recognizing Emotions in a Foreign Language. *J Nonverbal Behavior* 33:107-120. Doi 10.1007/s10919-008-0065-7

Pourtois, G., de Gelder, B., Vroomen, J., Rossion, B., Crommelinck, M. (2000). The time-course of intermodal binding between seeing and hearing affective information. *Neuroreport 11*(6):1329-33. PMID: 10817616.

Roberson, D., Davidoff, J. (2000). The categorical perception of colors and facial expressions: the effect of verbal interference. *Memory and Cognition 28*(6):977-86. PMID: 11105523.

Scherer, K.R., Banse, R., Wallbott, H.G. (2001). Emotion interences from vocal expressions correlate across languages and cultures. *J Cross-Cultural Psychology 32*(1):76-92. doi: 10.1177/0022022101032001009

Schröder, M. (2003). Experimental study of affect bursts. *Speech Communication* 40:99-116. doi.org/10.1016/S0167-6393(02)00078-X

Skinner, H.A. (1982). The Drug Abuse Screening Test. *Addictive Behavior 7*(4):363-371. PMID: 7183189

Tiddeman, B., Burt, D.M., Perrett, D. (2001). Prototyping and Transforming Facial Textures for Perception Research. *IEEE Computer Graphics and Applications 21*(5):42–50. doi: 10.1109/38.946630.

Tiddeman, B., Perrett, D. (2002). Transformation of Dynamic Facial Image Sequences Using Static 2D Prototypes. *The Visual Computer 18*(4): 218–25. doi:10.1007/s003710100142.

Tottenham, N., Tanaka, J.W., Leon, A.C., McCarry, T., Nurse, M., Hare, T.A., Marcus, D.J., Westerlund, A., Casey, B.J., Nelson, C. (2009). The NimStim set of facial expressions: judgments from untrained research participants. *Psychiatry Res*. *168*(3):242-9. PMID: 19564050.

Young, A. W., Rowland, D., Calder, A. J., Etcoff, N. L., Seth, A., Perrett, D. I. (1997). Facial expression megamix: Tests of dimensional and category accounts of emotion recognition. *Cognition*. 63:271-313. PMID: 926587

# CHAPTER 3 – PROSODIC INFLUENCE IN EMOTIONAL FACE PERCEPTION: EVIDENCE FROM FUNCTIONAL NEAR-INFRARED SPECTROSCOPY

## 3.1 Introduction

*3.1.1 Affective Faces and Voices.* Affective cues communicated by the face and voice of another individual are automatically integrated by the brain to form whole percepts of emotion. It has been suggested that all facial expressions can be categorized as belonging to one of six basic emotions (happiness, sadness, anger, surprise, fear, and disgust; Ekman & Friesen, 1976; Bartlett, Viola, Sejnowski, Golomb, Larsen, Hager, & Ekman, 1996), each of which is distinguished by the unique engagement of different sets of facial muscles (Ekman & Friesen, 1976). The vocal system appears to transmit a vocal analog of the six basic emotions via the coordination of multiple articulatory gestures which alter the rhythm, intensity, and intonation of a speaker's voice to create different prosodies (Schröder, 2003; Grandjean, et. al, 2005; Belin, Fillion-Bilodeau, & Gosselin, 2008; Juslin & Laukka, 2003; Patel, Scherer, Björkner, & Sundberg, 2011). Additionally, vocal and facial expressions of emotion may be inherently linked as affective vocalizations coincide with and are produced by the coordinated action of multiple vocal and facial muscles (Schröder, 2003; Belin, Fillion-Bilodeau, & Gosselin, 2008). Moreover, while these mechanisms may be innately linked at the physiological level, each modality appears to provide sufficient affective information to accurately recognize and identify emotion independently (Ekman & Friesen, 1976; Schröder 2003; Belin, Fillion-Bilodeau, & Gosselin, 2008). These findings indicate that emotion appears to be relatively well conserved across sensory modalities, allowing for multiple representations of the same internal state.

*3.1.2 Neural Substrates of Face Processing.* Faces are essential to communication and this fundamental role is reflected by their selective representation in the fusiform face area (FFA) in

the right hemisphere (Bowers, Bauer, Coslett, & Heilman, 1985; Kanwisher, McDermott, & Chun, 1997; Kanwisher & Yoval, 2006). Functional magnetic resonance imaging (fMRI) studies have shown that the FFA exhibits robust selectivity for faces, with responses occurring even when facial features are scrambled, re-arranged, inverted, or partly occluded (Kanwisher, McDermott, & Chun, 1997; Liu, 2003; Yovel & Kanwisher, 2004; Vuilleumier & Pourtois, 2007). The FFA appears to be engaged in a distributed network of brain areas spanning cortical and subcortical areas including the amygdala, posterior-temporal cortices, and ventral parietal cortices, as well as some portions of the somatosensory areas (Adolphs, 2002). Initially, invariant facial features are processed by the FFA, which possesses indirect connections to the posterior superior temporal sulcus (pSTS) and occipital face area (OFA) via the amygdala (Pessoa & Adolphs, 2010; Gauthier, et al., 2000; Rossion, et al., 2003; Winston, Vuilleumier, & Dolan, 2003; Haxby, Hoffman, & Gobbini, 2000; Kanwisher, McDermott, & Chun, 1997; Rotshtein, Henson, Treves, Driver, & Dolan, 2005; Fox, Young Moon, Iaria, & Barton, 2009; Schirmer & Adolphs, 2017). Further discrimination of facial features occurs in the OFA, which has been linked creating perceptual-representations of facial stimuli (Kanwisher, McDermott, & Chun, 1997; Haxby, Hoffman, & Gobbini, 2000; Winston, Vuilleumier, & Dolan, 2003; Rotshtein, et al., 2005; Fox et al., 2009; Leppänen & Nelson, 2009; De Winter, et al., 2015).

The pSTS is linked to perceiving the ‘changeable’ qualities of a person’s face, such as eye, lip, or cheek movements and plays a key role in audiovisual integration (Leppännen & Nelson, 2009; Haxby, Hoffman, & Gobbini, 2000; Winston, Vuilleumier, & Dolan, 2003; Rotshtein, et al., 2005; Fox et al., 2009; De Winter et al., 2015). Moreover, physically distinct, but overlapping areas of the STS respond to both the visual and auditory correlates of moving faces, with activations seen while viewing dynamic faces, listening to affective voices (Hoffman & Haxby,

2000; Yang, Rosenblau, Keifer, & Pelphrey, 2015), or during silent lip reading (Calvert, et al., 1997). Suggesting that these areas may possess a more holistic representation of emotion, which incorporates both the changeable aspects of faces with their concomitant vocalizations.

*3.1.3 Neural Substrates of Voice Processing.* Nonverbal affective voices have been shown to produce differential patterns of activation in spatially adjacent areas of the STS and STG of the right hemisphere (Belin, Zatorre, & Ahad, 2002; Von Kriegstein, Eger, Kleinschmidt, & Giraud, 2003; Kriegstein & Giraud, 2004; Ethofer, Van De Ville, Scherer, & Vuilleumier, 2009). These structures are arranged in a pattern homologous to that of language areas in the left hemisphere (Borod, et al., 2000; Borod et al., 2002; Adolphs, 2002; Belin, Zatorre, & Ahad, 2002; Belin, Fecteau, & Bédard, 2004; Gandour, Tong, Wong, Talavage, Dzemidzic, Xu, Li, & Lowe, 2004; Wildgruber, et al., 2004; Wildgruber, et al., 2005; Schirmer & Kotz, 2006; Kriegstein & Giraud, 2004; Wildgruber, Ackermann, Kreifelts, & Ethofer, 2006), with the progressive processing of emotional prosody evolving as activity moves from posterior to anterior portions of the STS/G (Belin, Zatorre, & Ahad, 2002; Belin, Fecteau, & Bédard, 2004; Schirmer & Kotz, 2006). Posterior aspects of the STS perform slow, low-level sensory analysis for the early discrimination and detection of emotion approximately 100 ms after stimulus presentation (Schirmer & Kotz, 2006; Wildgruber, et al., 2006). Subsequent attribution and integration with conceptual, emotion specific knowledge appears to first manifest in unique patterns of activation in the middle STG (Grandjean, et al., 2005; Wildgruber, et al., 2006; Johnstone, van Reekum, Oakes, & Davidson, 2006; Ethofer, Van De Ville, Scherer & Vuilleumier, 2009; Früholz & Grandjean, 2013; Kragel & LaBar, 2015) around 200 ms (Pourtois, et al., 2000; Johnstone, et al., 2006; Schirmer & Kotz, 2006). Evaluative judgments are made in the inferior frontal gyrus and orbital frontal cortex around 400 ms to be used in higher cognitive functions involved in emotional judgments (Schirmer & Kotz, 2006). This

processing stream occurs in parallel with other neural subsystems which process and access information about the speaker's identity, speech, and affective state (Belin et al., 2004). In this way, the STS may function as an auditory analog of the FFA, by exhibiting expertise in analyzing affective vocalizations to create a 'auditory faces' (Grandjean et al., 2005; Schirmer & Kotz, 2006; Schirmer & Adolphs, 2017).

Together, these findings indicate that emotion perception occurs via a distributed pattern of cortical and subcortical brain areas which appear to converge and engage partially overlapping regions in the pSTS/G (Adolphs, Damasio, Tranel, & Damasio, 1996; Wright et al., 2003; Adolphs, Tranel, & Damasio, 2003; Beauchamp, Argall, Bodurka, Duyn, & Martin, 2004; Kreifelts, Ethofer, Grodd, Erb, & Wildgruber, 2007; Ethofer, Bretscher, Wiethoff, Bisch, Schlipf, Wildgruber, & Kreifets, 2013; Schirmer & Adolphs, 2017). This functional convergence makes intuitive sense, as emotions are often conveyed via multiple perceptual channels, which can independently communicate emotion (Massaro & Egan, 1996; de Gelder & Vroomen, 2000). Regardless of the physical medium, presentation of a face, voice, or body movement impresses upon the observer the same perceptual experience. The neural correlate of this supramodal representation appears to occur in the pSTS, wherein patterns of activity change between emotions, but not between different modalities within an emotion (Kreifelts et al., 2007; Wright et al., 2003; Kreifelts, 2009; Ethofer, et al., 2013; Peele, Atkinson, & Vuilleumier, 2010; Schirmer & Adolphs, 2017). Additionally, simultaneous presentation of affective voices and faces elicited activity in areas identical to those when voices and faces were presented separately (Kreifelts et al., 2007). This overlap may represent the common engagement of several neural structures which support the perception, experience, and expression of emotion (Davidson, 1995).

*3.1.4 Limitations of Existing Neuroimaging Findings.* While the number of multimodal imaging studies has steadily increased (Klasen, Chen, & Mathiak, 2012), most findings are based upon studies of unimodal faces (de Gelder & Vroomen, 2000; Gerdes, Wieser, Bublatzky, Kusay, Plichta, & Alpers, 2013; Gerdes, Wieser, & Alpers, 2014). Fewer studies incorporate affective vocalizations (de Gelder & Vroomen, 2000; Belin, Fillion-Bilodeau, & Gosselin, 2008; Korb, Frühholz, & Grandjean, 2015). This disparity may be related to the assumed similarity between the processing of emotional visual and auditory stimuli (King & Nelken, 2009), with faces serving as a 'prototype' of how affective expressions in other modalities should be processed (King & Nelken, 2009; Schirmer & Adolphs, 2017). Such speculation undermines the complex conceptual and linguistic information carried by a speaker's voice (Borod, et al. 2000). This is of critical importance in the interpretation of multimodal studies as verbal stimuli may unintentionally activate language areas which are unrelated to the emotion of interest (Belin, Zatorre, & Ahad, 2002), and it is noteworthy that verbal or semantic vocalizations do result in different patterns of neural activity than nonverbal prosodic voices (Belin, Zatorre, & Ahad, 2002; Schirmer & Kotz, 2006; Frühholz & Grandjean, 2013). This division is striking as paralinguistic, nonverbal vocalizations selectively exhibit activity in the temporal lobe of the right hemisphere, while verbal utterances are associated with bilateral activity in language areas (Schirmer & Kotz, 2006; Scherer & Adolphs, 2017). Additionally, a failure to test vocal and facial stimuli as independent components may have made these multimodal findings more difficult to interpret, as there was no quantitative account of how each modality was differentially processed when presented alone and then integrated together (Borod, et al. 2000).

*3.1.5 Current Study.* The current study examined unimodal and bimodal visual and auditory components of affect perception using static face images and nonverbal vocal utterances. Using a

region-of-interest approach, functional near-infrared spectroscopy (fNIRS) was used to quantify changes in oxygenated-hemoglobin (HbO) levels in bilateral temporoparietal junction (TPJ) areas. These areas are associated with the integration of bimodal social stimuli (right TPJ) and perspective taking (left TPJ; Decety & Lamm, 2007; Samson, Apperly, Chiavarino, & Humphreys, 2004). We predicted that the prosody of the simultaneously presented voice would bias the participant's perception of the stimulus and that this would be evidenced by increased HbO levels in bilateral TPJs to both angry and happy face and voice combinations compared with neutral voicings paired with faces.

## 3.2 Materials and Methods

*3.2.1 Participants*. Thirty-nine subjects were drawn from the undergraduate introductory psychology subject pool volunteers from Colorado State University (Table 3.1). The protocol was approved by the Colorado State University Institutional Review Board and all participants provided informed consent before taking part in the procedures.

**Table 3.1** *Subject demographics*

| *n* | Age | Gender | |
|---|---|---|---|
| | *M*(*SD*) | M | F |
| 39 | 20.37(1.19) | 17 | 23 |

Exclusionary criteria were based on self-report and included: past or present neurological or psychiatric diagnosis, history of developmental disability, traumatic brain injury, current tobacco use, neurological disorders, visual acuity of 20/20 or worse with or without correction, and chronic or current substance abuse within the past three months.

3.2.2 *Auditory Stimuli.* Auditory stimuli were obtained from the Montreal Affective Voices database (Belin, Fillion-Bilodeau, & Gosselin, 2008), in which professional actors produced short, nonverbal affective interjections of the vowel /a/, which sounds similar to the a in "ah" in spoken English. The current stimuli were chosen because they effectively convey emotion, they are not synthetic, and are free of semantic or linguistic information that may indirectly bias participants'

responses. Three vocalizations expressed in angry, happy, and neutral prosody were chosen for each actor (1 male and 1 female) resulting in a total of six vocalizations. These stimuli have previously been matched and validated for valence (negative, positive), arousal, and perceived intensity (Belin, Fillion-Bilodeau, & Gosselin, 2008). All vocal stimuli were 993 ms in length. However, it should be emphasized that identification of anger and happiness in vocal stimuli appears to occur on a similar time course when compared to other emotions (Pell & Kotz, 2011), which is imperative to imaging techniques with excellent temporal precision.

*3.2.3 Facial Stimuli*. Face stimuli were obtained from the NimStim database (Tottenham, Tanaka, Leon, McCarry, Nurse, Hare, Marcus, Westerlund, Casey, & Nelson, 2009). This dataset used professional actors with natural hair and makeup. Two images (one angry and one happy closed-mouth image from each actor) were selected from a subset of 20 actors (10 men) from the NimStim database, for a total of 40 face images. Images were converted to grayscale and cropped tightly around the face so that no hair, neck or clothing was visible. Twenty emotional continua were generated, one for each actor, using Psychomorph software (Tiddeman, Burt, & Perrett, 2001; Tiddeman & Perrett, 2002). Each continuum consisted of two end-point prototype images (angry or happy), which were morphed together in seven steps (two endpoints and 5 morphs, in 12.5% steps) so that the mid-point image would be a 50% combination of each prototype image. Individual face templates were created for each end-point image using 182 manually placed points. Faces with closed mouths were selected to facilitate morphing.

*3.2.4 NIRS Paradigm*. Participants were presented with three classes of stimuli: face+voice (FV), voice only (VO), and face only (FO). These stimuli were used to create seven conditions, one for each prosody (happy, angry, neutral) for the FV and VO conditions with one condition for the FO stimuli (Figure 3.1, top).

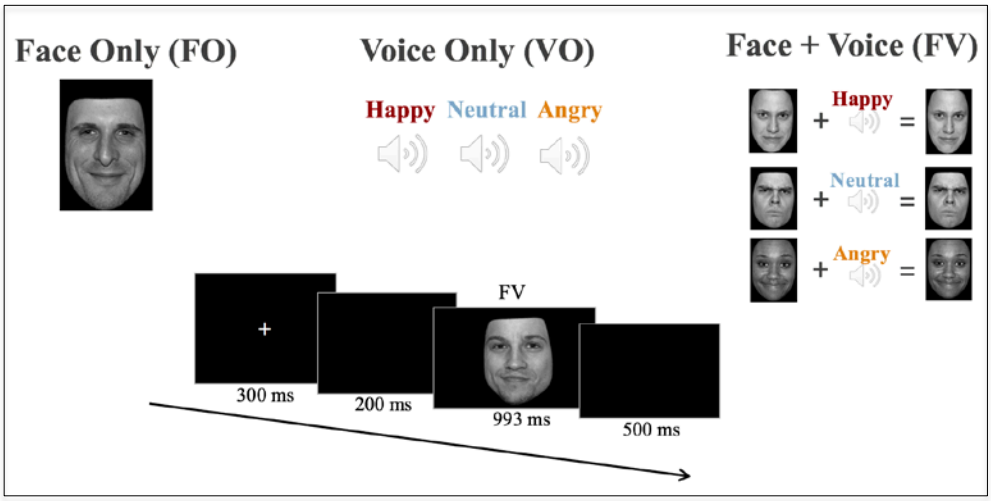


**Figure 3.1:** (Top) Examples of face, voice, and face and voice stimuli. (Bottom) Example of a single trial with a face + voice (FV) stimulus.

Binaural auditory stimulation (70 dB SPL) was delivered via EAR 3a foam insert earphones. Morphed face stimuli were presented alone or simultaneously with auditory stimuli on an LED monitor at 240 Hz refresh rate located 45 cm in front of the subject. Face stimuli subtended 7.62 degrees of vertical visual angle and 5.72 degrees of horizontal angle. Each trial began with a white fixation cross on a black background for 300 ms, followed by a 200 ms pause, after which a voice-only (VO), face-only (FO), or face+voice (FV) stimulus was presented for a duration of 993 ms, followed by a blank black screen for 500 ms, for a total trial time of 1993 ms (Figure 3.1, bottom). Subjects were instructed to identify the emotion expressed by the actor in each trial in a 2-alternative (happy/not happy), forced choice procedure on a Cedrus RB-730 (Cedrus Corporation, United States) response pad without specific reference to the face or voice.

Stimuli were presented in a hybrid block design presented in E-Prime 2 Professional (Psychology Software Tools, Inc., United States). The experiment contained a total of 980 trials (20 actors x 7 conditions x 7 faces on a continuum) with trials grouped into pseudo-random VO,

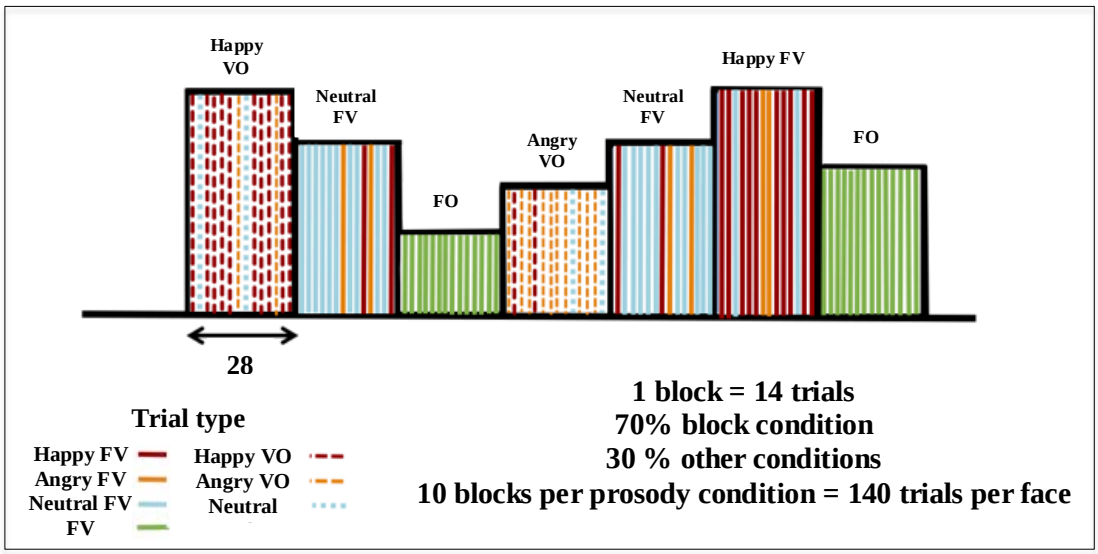


**Figure 3.2:** Schematic of trial and block organization.

FO, and FV condition blocks (Figure 3.2). Blocks were defined by their stimulus type (FO, FV, VO) and condition (happy FV, angry FV, neutral FV, happy VO, angry VO, neutral VO, FO). All blocks contained 14 trials. Block condition was indicated by the trial type that was in the majority. For the three FV and VO conditions, approximately 70% of the trials were the same as the block voice condition, and the remaining 30% was divided equally between the two remaining voice conditions. Each condition was shown in 10 blocks, for a total of 70 blocks. Face and voice gender were matched for all FV conditions. Two resting periods of 2-minute duration were added at the beginning and middle of the experiment where subjects were instructed to relax, while a central fixation cross was shown on the screen. The total duration of the scan was approximately 36.5 minutes.

*3.2.5 NIRS Instrumentation*. Diffuse optical data was acquired using a continuous wave NIRScoutX (NIRScout; NIRx Medical Technologies, Los Angeles, CA, USA) NIRS system

which can record from up to 32 multiplexed silicon dioxide photodetectors. In our montage, 16 detectors were located over each hemisphere. The optode array contained 28 source positions (light emitting diodes) operating at two wavelengths 760 and 850 nm. Data were acquired at a sampling frequency of 3.92 Hz. Sources and detectors were manually inserted into special NIRS recording caps (Brain Products GmbH, Germany; Easycap GmbH, Germany) configured in a standard 10-05 International Electrode system manner (Easycap montage M15). This arrangement distributed sources and detectors so that they were located approximately 3 cm apart, to produce a total of 105 channels, in an attempt to maximize coverage of the cortical surface and to obtain high-resolution estimates of chromophore concentrations (Scholkmann, et al., 2014).

*3.2.6 NIRS Data Analysis.* Recordings were analyzed in the spm_fnirs software package for Matlab (Tak, Uga, Flandin, Dan, & Penny, 2016), where data were cleaned of motion artifact (Scholkmann, Spichtig, Muehlemann, & Wolf, 2010), high pass filtered at 0.01 Hz, and temporally smoothed with a 5.0 s moving window to reduce cardiac and respiration noise. Data were compared using nine contrasts (angry prosody>neutral prosody, happy prosody>neutral prosody, angry prosody>neutral prosody, angry FV>neutral FV, happy FV>neutral FV, angry FV>happy FV, all voices > Face only, all face and voices > Face, all FV> all voices) All NIRS channels were first analyzed using a whole-brain approach, by implementing a general linear model design matrix to perform first-level statistics on HbO data. Second-level statistics were performed on all resulting contrasts (p-value .05) to reveal any significant channels. Multiple comparisons were corrected by FDR set at q = .05.

## 3.3 Results

Three optodes (two emitters and one detector) were eliminated from all scans as they failed quality control in more than half of the subjects' datasets. A gain check was used to assess

individual channel quality – channels with gains higher than 6 but otherwise passing quality controls were interpolated to form 103 channels to ensure that all subjects shared a common set of channels for comparisons.

*3.3.1 Face and Voice Contrasts.* While both the angry and happy FV conditions exhibited greater concentrations of HbO when compared with the neutral FV condition, the resulting patterns of activity were distinctly different (Figure 3.3). The angry-neutral FV contrast revealed a pattern of increased HbO that was primarily restricted to motor areas in right hemisphere (Figure 3.3b). Such activations spanned prefrontal areas, with increased HbO occurring in superior and inferior aspects of the precentral gyrus. A handful of channels also exhibited significantly greater activity when compared to the neutral FV condition: supramarginal gyrus, middle superior temporal gyrus (STG), inferior postcentral gyrus, and the inferior occipital gyrus (IOG). The happy-neutral FV contrast exhibited a strikingly different pattern of activity, with increased levels of HbO occurring bilaterally (Figure 3.3a). Activations in the right hemisphere spanned posterior dorsal areas which included superior parietal lobules (SPL), superior, middle, and inferior occipital gyri. For all contrasts, information regarding the NIRS channel number, anatomical locations, Montreal Neurological Institute (MNI) coordinates, and corresponding EEG locations are shown in tables directly after each contrast.

Bilateral activations appeared in left and right Temporoparietal junctions (TPJ), as well as posterior and middle sections of the STG. In the left hemisphere activations appeared in prefrontal and precentral gyrus areas, inferior frontal gyrus (IFG), superior frontal gyrus (SFG), and TPJ. When compared against one another, the angry FV condition exhibited greater HbO in right inferior precentral gyrus than the happy FV condition (Figure 3.3c). Interestingly, the happy FV

condition exhibited greater HbO in the right superior occipital gyrus, left dorsolateral prefrontal areas, and superior postcentral gyrus than the angry FV condition.

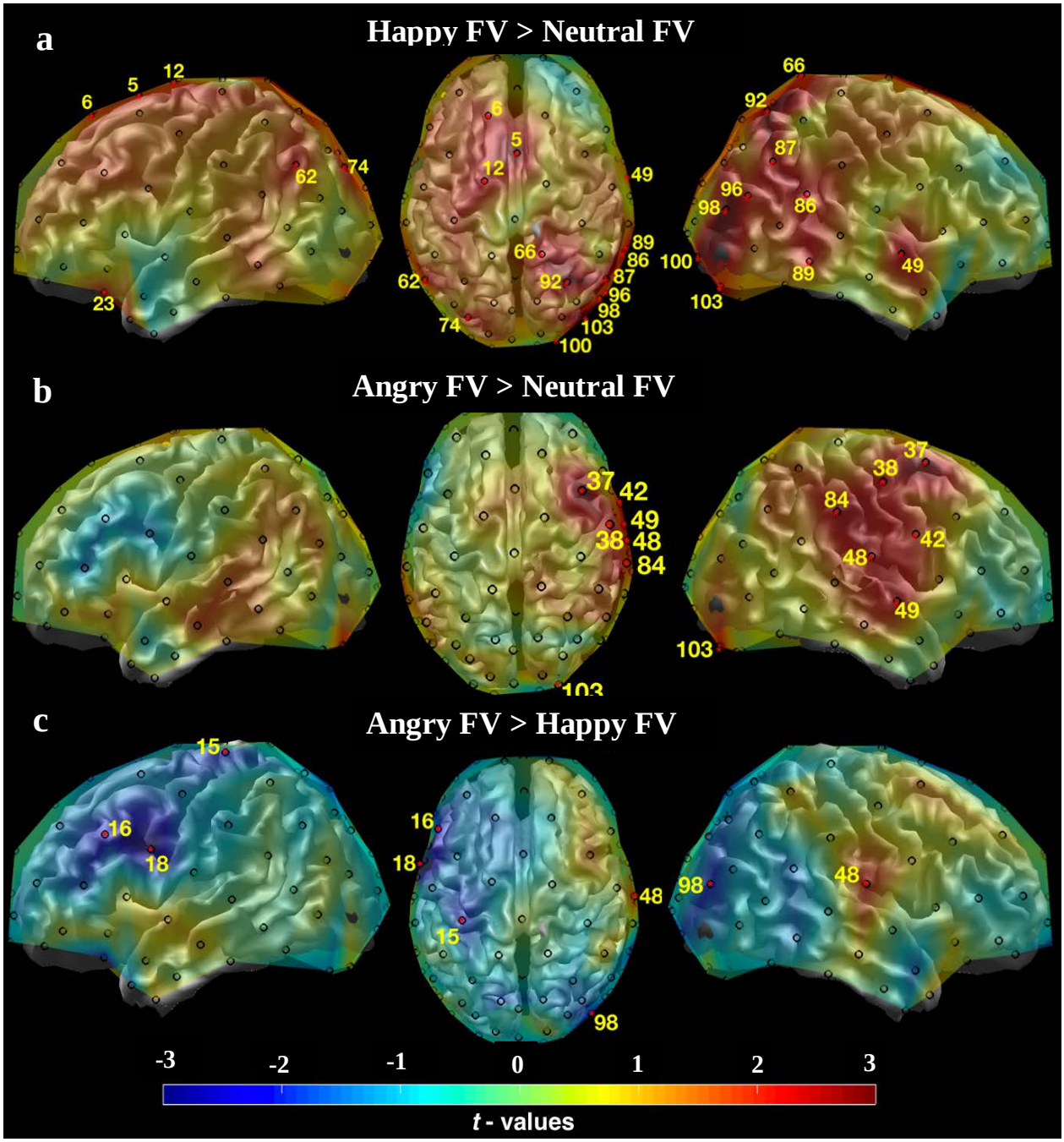


**Figure 3.3:** Results of the bimodal (face+voice (FV)) condition contrasts. Areas with significant differences in HbO concentrations are indicated by red markers, with the associated channel number shown in yellow. The magnitude and direction of *t*-values is represented by the color bar shown on the bottom of the figure. All channels survived FDR correction at q = .05, all significant at $p < .05$.

When compared to the FO condition, both the angry and happy FV conditions exhibited activity in the right hemisphere, with a similar anterior-posterior division as the previous FV-FV contrasts (Figure 3.4). Activity included, but was not limited to the TPJ for both contrasts. The happy FV condition exhibited increased HbO activity in posterior occipital and dorsal association areas (Figure 3.4a). Activity for the angry FV condition was primarily localized to motor areas (Figure 3.4b). Additionally, while greatly reduced, a similar increase in right lateralized anterior-posterior HbO activity appeared when both the angry and happy FV conditions were compared with their respective VO conditions (Figure 3.5). Happy FV activity appeared in right TPJ and posterior occipital areas. The angry FV exhibited increased HbO in frontal motor areas, but showed decreased HbO in anterior portions of the right temporal lobe (Figure 3.5b).

**Table 3.2** *Significant Channels and Anatomical Locations for All Face-Voice (FV) Contrasts.*

| Brain Area | NIRS channel | MNI Coordinates | | | EEG location | *t* value |
|---|---|---|---|---|---|---|
| | | x | y | z | | |
| ***Happy FV > Neutral FV*** | | | | | | |
| Bilateral supplementary motor area | 5 | 0.7 | 9.3 | 68.7 | FCz – Cz | 2.94 |
| L Supplemental motor area | 6 | -17.0 | 31.7 | 60.0 | FCz – FC1 | 2.11 |
| L Superior frontal gyrus | 12 | -19.0 | -7.7 | 75.3 | C1 – Cz | 2.37 |
| L Inferior frontal gyrus | 23 | -44.3 | 25.0 | -22.7 | F5 – F7 | 2.03 |
| R Middle superior temporal gyrus | 49 | 69.0 | -4.7 | -4.3 | C6 – FC6 | 2.65 |
| L Temporoparietal junction | 62 | -55.7 | -67.0 | 37.3 | CP3 – P3 | 2.02 |
| R Superior parietal lobule | 66 | 16.0 | -51.7 | 77.0 | CPz – CP2 | 2.43 |
| L Superior occipital gyrus | 74 | -29.3 | -89.7 | 35.7 | P1 – PO1 | 2.19 |
| R Posterior superior temporal sulcus | 86 | 67.0 | -50.3 | 23.7 | CP4 – CP6 | 2.45 |
| R Temporoparietal junction | 87 | 55.7 | -66.3 | 38.7 | CP4 – P4 | 2.41 |
| R Occipitotemporal junction | 89 | 69.0 | -49.0 | -8.0 | TP8 – CP6 | 2.19 |
| R Posterior parietal cortex | 92 | 31.7 | -68.3 | 61.3 | P2 – CP2 | 3.12 |
| R Inferior temporoparietal junction | 96 | 51.7 | -78.3 | 22.3 | P6 – P4 | 2.10 |
| R Superior occipital gyrus | 98 | 43.3 | -88.7 | 15.3 | P6 – PO2 | 3.34 |
| R Middle occipital gyrus | 100 | 26.7 | -102.3 | -8.3 | PO4 – PO6 | 2.72 |
| ***Angry FV> Neutral FV*** | | | | | | |
| R Superior frontal gyrus | 37 | 42.3 | 8.7 | 60.7 | C2 – FC2 | 3.13 |
| R Superior precentral gyrus | 38 | 59.0 | -11.3 | 51.7 | C2 – C4 | 2.24 |
| R Inferior precentral gyrus | 42 | 66.3 | 4.3 | 27.0 | FC4 – C4 | 2.04 |
| R Inferior postcentral gyrus | 48 | 70.0 | -16.7 | 16.7 | C6 – C4 | 2.31 |
| R Middle superior temporal gyrus | 49 | 69.0 | -4.7 | -4.3 | C6 – FC6 | 2.42 |
| R Supramarginal gyrus | 84 | 69.0 | -33.3 | 37.7 | CP4 – C4 | 2.48 |
| R Inferior occipital gyrus | 103 | 33.0 | -90.0 | -26.0 | PO10 – PO6 | 2.09 |
| ***Angry FV > Happy FV*** | | | | | | |
| L Superior postcentral gyrus | 15 | -37.7 | -31.3 | 71.0 | C1 – CP1 | -2.17 |
| L Dorsolateral prefrontal cortex | 16 | -52.7 | 24.7 | 33.3 | FC3 – FC1 | -2.39 |
| L Inferior precentral gyrus | 18 | -64.0 | 3.3 | 26.3 | FC3 – C3 | -3.15 |
| R Inferior postcentral gyrus | 48 | 70.0 | -16.7 | 16.7 | C6 – C4 | 2.23 |
| R Superior occipital gyrus | 98 | 43.3 | -88.7 | 15.3 | P6 – PO2 | -2.13 |

Note: NIRS channel locations are shown in MNI coordinates. Source-detector pairs are given according to 10/05 EEG electrode positions. Sources are shown on the left of each electrode pair. All channels significant at $p < .05$, corrected.

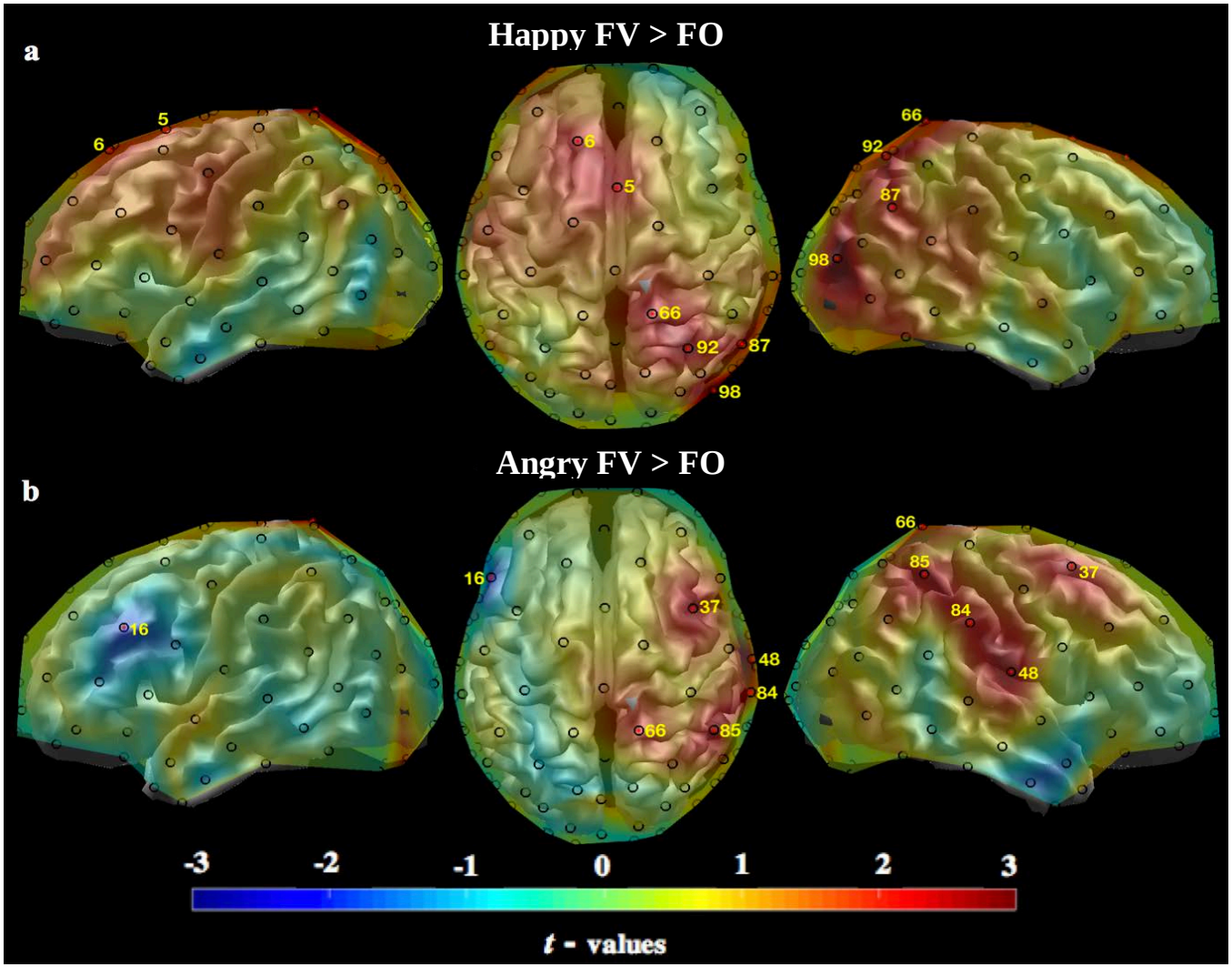


**Figure 3.4:** Results of the bimodal (face+voice (FV)) face only (FO) condition contrasts. Areas with significant differences in HbO concentrations are indicated by red markers, with the associated channel number shown in yellow. The magnitude and direction of *t*-values is represented by the color bar shown on the bottom of the figure. All channels survived FDR correction at q = .05, all significant at $p < .05$.

**Table 3.3** *Significant Channels and Anatomical Locations for All Face-Voice (FV) Face (FO) Contrasts.*

| Contrast | NIRS channel | MNI Coordinates | | | EEG location | *t* value |
|---|---|---|---|---|---|---|
| | | x | y | z | | |
| ***Happy FV > Face*** | | | | | | |
| Bilateral supplementary motor area | 5 | 0.7 | 9.3 | 68.7 | FCz – Cz | 2.34 |
| L Supplemental motor area | 6 | -17.0 | 31.7 | 60.0 | FCz – FC1 | 2.41 |
| R Superior parietal lobule | 66 | 16.0 | -51.7 | 77.0 | CPz – CP2 | 2.47 |
| R Temporoparietal junction | 87 | 55.7 | -66.3 | 38.7 | CP4 – P4 | 2.15 |
| R Posterior parietal cortex | 92 | 31.7 | -68.3 | 61.3 | P2 – CP2 | 2.70 |
| R Superior occipital gyrus | 98 | 43.3 | -88.7 | 15.3 | P6 – PO2 | 3.45 |
| ***Angry FV > Face*** | | | | | | |
| L Dorsolateral prefrontal cortex | 16 | -52.7 | 24.7 | 33.3 | FC3 – FC1 | -2.08 |
| R Superior frontal gyrus | 37 | 42.3 | 8.7 | 60.7 | C2 – FC2 | 2.69 |
| R Inferior postcentral gyrus | 48 | 70.0 | -16.7 | 16.7 | C6 – C4 | 2.84 |
| R Superior parietal lobule | 66 | 16.0 | -51.7 | 77.0 | CPz – CP2 | 2.21 |
| R Supramarginal gyrus | 84 | 69.0 | -33.3 | 37.7 | CP4 – C4 | 2.25 |
| R Superior parietal lobule | 85 | 51.3 | -51.7 | 57.7 | CP4 – CP6 | 2.67 |

Note: NIRS channel locations are shown in MNI coordinates. Source-detector pairs are given according to 10/05 EEG electrode positions. Sources are shown on the left of each electrode pair. All channels significant at $p < .05$, corrected.

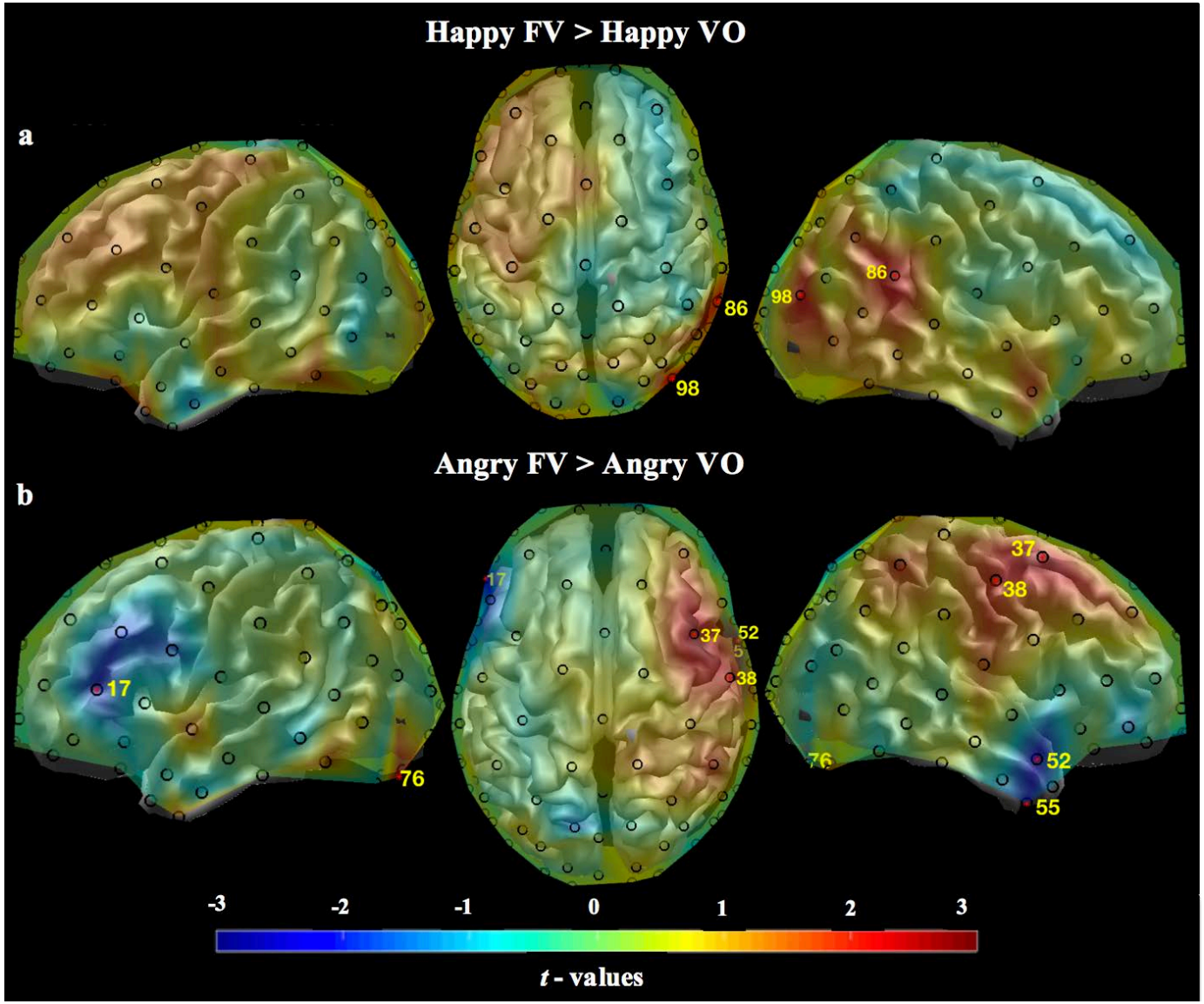


**Figure 3.5:** Results of the bimodal (face+voice (FV)) voice only (VO) condition contrasts. Areas with significant differences in HbO concentrations are indicated by red markers, with the associated channel number shown in yellow. All channels survived FDR correction at q = .05, and are significant at $p < .05$.

**Table 3.4** *Significant Channels and Anatomical Locations for All Face-Voice (FV) Voice Only (VO) Contrasts*

| Contrast | NIRS channel | MNI Coordinates | | | EEG location | *t* value |
|---|---|---|---|---|---|---|
| | | x | y | z | | |
| ***Happy FV > Happy VO*** | | | | | | |
| R Supramarginal gyrus | 86 | 67.0 | -50.3 | 23.7 | CP4 – P4 | 2.55 |
| R Superior occipital gyrus | 98 | 43.3 | -88.7 | 15.3 | P6 – PO2 | 2.37 |
| ***Angry FV > Angry VO*** | | | | | | |
| L Dorsolateral prefrontal gyrus | 17 | -55 | 34.7 | 10 | FC5 – FC3 | -2.17 |
| R Superior frontal gyrus | 37 | 42.3 | 8.7 | 60.7 | C2 – FC2 | 2.75 |
| R Precentral gyrus | 38 | 59.0 | -11.3 | 51.7 | C2 – C4 | 2.12 |
| R Anterior middle temporal gyrus | 52 | 63.0 | 6.0 | -23.0 | FT8 – FC6 | -2.42 |
| R Inferior anterior temporal gyrus | 55 | 55.7 | 1.7 | -41.7 | FT10 – F10 | -2.04 |

*Note*. NIRS channel locations are shown in MNI coordinates. Source-detector pairs are given according to 10/05 EEG electrode positions. Sources are shown on the left of each electrode pair. All channels significant at $p < .05$, corrected.

*3.3.2 Voice Only Contrasts.* When contrasted with the neutral voice condition, both the angry and happy voice conditions exhibited decreased activity bilaterally (Figure 3.6a,b). Similar to the activations witnessed in the bimodal contrasts, these patterns were not overlapping, with angry voice showing lower concentrations of HbO in the left IOG and right precentral gyrus (Figure 3.6b). Whereas, the happy voice condition exhibited decreased HbO in left MTG and posterior areas of the right STG (Figure 3.6a). When compared against one another, significant differences were right lateralized, with angry voice showing greater concentrations of HbO than the happy voice condition in the anterior portion of the temporal lobe, and IFG (Figure 3.6c).

*3.3.3 Face Only Contrasts.* The summed activity of the FV conditions was contrasted with the face only condition which revealed that the face only condition produced greater activations in right SPL, and inferior aspects of the occipital-temporal junction (Figure 3.7).

Interestingly, the face condition exhibited greater HbO in the anterior of the temporal lobe when compared to the bimodal conditions. Collapsed across voice conditions, emotional voices exhibited more HbO bilaterally in superior portions of the postcentral gyrus and right lateralized activity in the SPL.

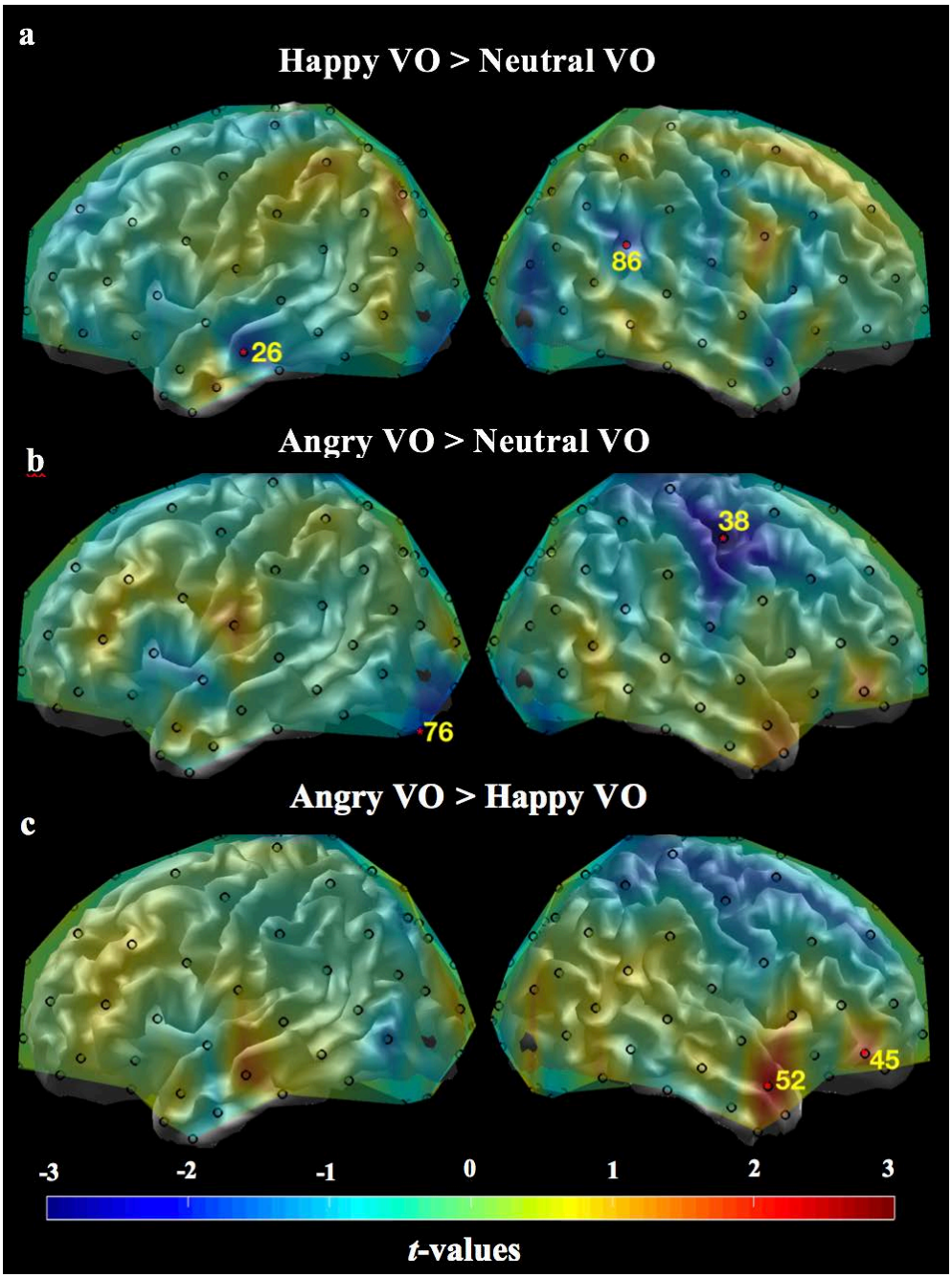


**Figure 3.6:** Map of significant *t*-values for the voice only (VO) contrasts. The magnitude and direction of *t*-values is represented by the color bar shown on the bottom of the figure. All channels survived FDR correction at $q < .05$.

**Table 3.5** *Significant Channels and Anatomical Locations for All Voice Only (VO) Contrasts.*

| Brain Area | NIRS channel | MNI Coordinates | | | EEG location | *t* value |
|---|---|---|---|---|---|---|
| | | x | y | z | | |
| ***Happy VO > Neutral VO*** | | | | | | |
| L Middle temporal gyrus | 26 | -70.0 | -19.7 | -17.3 | C5 – T7 | -2.10 |
| R Supramarginal gyrus | 86 | 67.0 | -50.3 | 23.7 | CP4 – P4 | -2.05 |
| ***Angry VO > Neutral VO*** | | | | | | |
| R Precentral gyrus | 38 | 59.0 | -11.3 | 51.7 | C2 – C4 | -3.14 |
| L Inferior occipital gyrus | 76 | -33.7 | -89.3 | -26.3 | PO9 – PO5 | -2.35 |
| ***Angry VO > Happy VO*** | | | | | | |
| R Inferior frontal gyrus | 45 | 53.0 | 44.7 | -10.0 | F6 – F4 | 2.11 |
| R Anterior middle temporal gyrus | 52 | 63.0 | 6.0 | -23.0 | FT8 – FC6 | 2.58 |

*Note*. NIRS channel locations are shown in MNI coordinates. Source-detector pairs are given according to 10/05 EEG electrode positions. Sources are shown on the left of each electrode pair. All channels significant at $p < .05$, corrected.

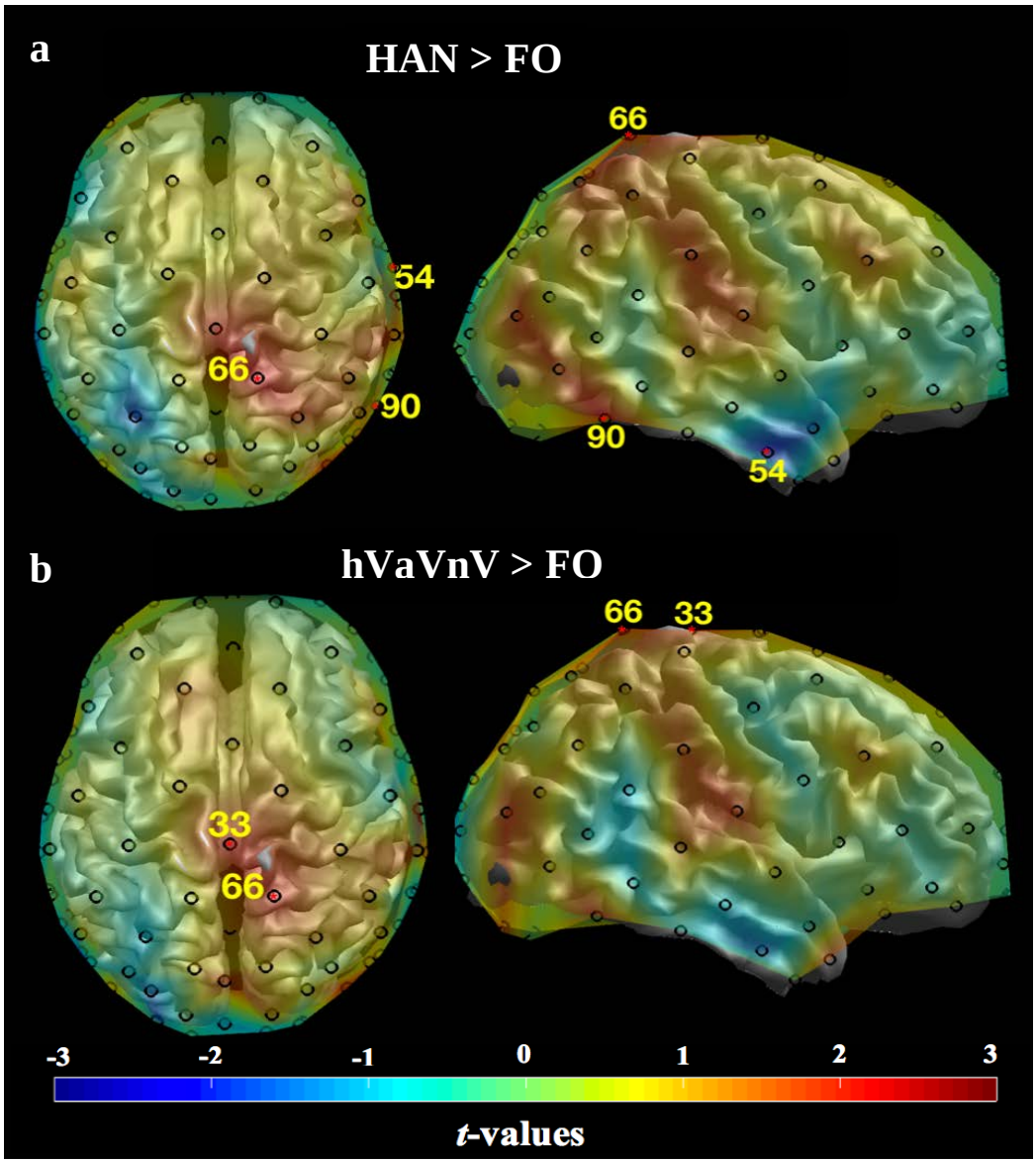


**Figure 3.7:** Results of the bimodal (face+voice (FV)) and unimodal (voice only) face only (FO) contrasts, collapsed across prosody conditions. Areas with significant differences in HbO concentrations are indicated by red markers, with the associated channel number shown in yellow. All channels survived FDR correction at q = .05, and are significant at $p < .05$.

**Table 3.6** *Significant Channels and Anatomical Locations for All Face Only (FO) Contrasts.*

| Contrast | NIRS channel | MNI Coordinates | | | EEG location | *t* value |
|---|---|---|---|---|---|---|
| | | x | y | z | | |
| ***HAN > FO*** | | | | | | |
| R Inferior temporal gyrus | 54 | 64.0 | -8.7 | -31.3 | FT8 –T8 | -2.18 |
| R Superior parietal lobule | 66 | 16.0 | -51.7 | 77.0 | CPz – CP2 | 2.22 |
| R inferior occipitotemporal junction | 90 | 61.0 | -60.7 | -19.7 | TP8 – P8 | 2.22 |
| ***HAN Voice > FO*** | | | | | | |
| Bilateral superior precentral gyrus | 63 | -0.3 | -30.7 | 76.0 | CPz – Cz | 2.06 |
| R Superior parietal lobule | 66 | 16.0 | -51.7 | 77.0 | CPz – CP2 | 2.06 |

Note: Collapsed across happy, angry, and neutral face+voice conditions (HAN), collapsed across voice conditions (HAN Voice). All channels significant at $p < .05$, corrected.

## 3.4 Discussion

The current study evaluated oxygenated hemoglobin (HbO) levels in bilateral parietal-occipital areas as participants were presented with images of affective facial expressions and vocal utterances voiced in a happy, angry, or neutral prosody. This design enabled the comparison of both the unimodal (face or voice presented alone) and bimodal (face and voice (FV)) components of affect expression. We hypothesized that HbO activity for the bimodal conditions would vary by prosody, with the angry and happy FV conditions exhibiting higher HbO levels in bilateral temporoparietal junctions (TPJ) than the neutral FV condition, and the angry FV condition showing greater HbO activity in these areas when compared to the happy FV condition. The data showed that the happy FV condition exhibited significantly greater HbO activity in these regions when compared to both the angry and neutral FV conditions (Figure 3.3a,c). While the happy FV exhibited greater HbO activity in bilateral TPJs than the neutral FV condition, differences in TPJ activity between the happy and angry FV conditions only appeared in the right hemisphere. These results are intriguing as the left and right TPJs are associated with different aspects of social communication, with integration of vocal and facial social stimuli occurring in the right TPJ and perspective taking being localized to the left TPJ. This suggests that TPJ activity may be linked to the valence of the prosodic voice, as both conditions contained the same faces, but only differed in the prosody of the voice. Moreover, the happy FV condition may have selectively engaged this area as there was no difference in TPJ activity between the neutral and angry FV conditions (Figure 3.3b).

Outside of these regions of interest, the happy and angry FV conditions exhibited pronounced differences in patterns of activity in the right hemisphere, with a distinct anterior-posterior division appearing between emotions (Figure 3.3a,b). Both the anterior and posterior

regions of the right hemisphere are essential to multimodal affect perception, as damage to either area may result in deficits in identifying, recognizing, or producing emotional expressions (Yuvaraj, Murugappan, Norlinah, Sundaraj, & Khairiyah, 2013). The right hemisphere is thus generally involved in distinguishing and associating vocal and facial information with their relevant conceptual representations. Additionally, these behavioral sequelae may be connected to the fundamental role of the right hemisphere in processing paralinguistic information related to a speaker's age, gender, or emotional state (Belin, Fecteau, Bedard, 2004; Schirmer & Kotz, 2006). Prosody, is thus an auditory extension of a speaker's identity and is thought to have a more pronounced effect on emotion perception than other sensory modalities (Brancucci, Lucci, Mazzantenta, & Tommasi, 2009). The impact of prosody in affect perception is underscored by the ability of affective voices to bias the perception of emotional facial expressions in the direction of the simultaneously presented prosody (de Gelder & Vroomen, 2000). Importantly, this effect is not tied to stimulus valence as both negative and positively valenced stimuli were shown to bias face perception (de Gelder & Vroomen, 2000). Together, these findings show that activity in the right hemisphere may represent a more general role in speaker identification and that this identification may be mediated by a speaker's prosody.

Results from the current study support and extend this assertion as activity was lateralized to the right hemisphere and seemingly subdivided by valence into two anterior and posterior subdivisions (Figure 3.3a,b). A similar functional division has been reported in the stroke literature, where patients sustaining damage to anterior regions in the right hemisphere exhibited deficits in recognizing negatively valenced emotions, but this effect was not seen in patients with posterior damage (Harciarek & Heilman, 2009). These findings complement those of the current study, which found that the angry FV condition exhibited increased HbO in the right hemisphere,

with activations primarily localized to right frontal and somatomotor areas (Figure 3.3b). Additionally, findings from lesion and traumatic brain injury studies have solidified the role of the right hemisphere in anger mimicry and perception, as individuals with right hemisphere damage can mimic and label happy, but not angry facial expressions (Adolphs, et al. 2000; Adolphs, 2002; Bailey, Henry, & Varcin, 2012). Further, this segregation to frontal and motor regions may represent an inherent difference in responding to and processing highly arousing, negatively valenced facial expressions (Blair & Cipolotti, 2000; Ongür & Price, 2000; Adolphs, 2002). Collectively, these findings indicate that frontal regions in the right hemisphere may possess a special motor representation for angry stimuli.

In contrast to anger, a homologous posterior anatomical region has not been reported for positively valenced emotions (Adolphs, Jansari, & Tranel, 2001; Harciarek & Heilman, 2009). Rather, happiness appears to be represented in several cortical and subcortical areas, with bilateral activity appearing in somatosensory and posterior association areas (Adolphs, Jansari, & Tranel, 2001; Amodio & Frith, 2006; Keysers & Gazzola, 2007). The happy FV condition displayed a similar pattern of posterior activity in right superior parietal and lateral occipital regions, with bilateral activity appearing in the TPJs. Again, while this lateralization may reflect the perceptual weight that vocalizations carry in affect perception, the posterior segregation of activity highlights the involvement of brain areas associated with visualizing, responding to and mentalizing another individual's psychological state (Amodio & Frith, 2006; Keysers & Gazzola, 2007). Additionally, these findings may reflect the automatic mimicry and approach behaviors evoked by happy faces (Fridlund, 1991; Hess & Fischer, 2014; Seidel, et al., 2010).

When compared against one another, the happy FV stimuli exhibited greater HbO activity over the right occipital region than the angry FV stimuli (Figure 3.3c). Most interestingly, the

happy FV condition exhibited greater levels of HbO in left dorsolateral prefrontal cortex (DLPFC). These data provide indirect support for the role of the right hemisphere in selectively processing angry facial expressions (Adolphs, et al. 2000; Adolphs, 2002; Adolphs, Tranel, & Damasio, 2003; Bailey, Henry, & Varcin, 2012). However, neither hemisphere displays a similar specialization for happy faces (Adolphs, et al. 2000; Adolphs, 2002). These findings may dually reflect the unilateral specialization of the right hemisphere in processing angry stimuli, as well as the more general, bilateral activity evoked by happy stimuli.

What should be emphasized is that despite their spatial segregation, the angry and happy FV conditions exhibited overlapping areas of increased HbO in the middle superior temporal gyrus (mSTG) and the right occipital region (ROR), areas that are essential to the fine discrimination and initial integration of affective vocal and facial cues with their corresponding emotion-specific perceptual-representations (Köchel, et al., 2011; Capilla, Belin, & Gross, 2013; Grandjean, et al., 2005). Further, these activations were only present when the angry and happy FV conditions were compared to the neutral FV condition, but not in direct comparisons between emotion conditions (Figure 3.3c), suggesting that while activity in mSTG and ROR may be mediated by emotional valence, the regions do not appear to be emotion specific. This is consistent with other neuroimaging studies, which have indicated that the ROR and mSTG are sensitive to the physical attributes of affective stimuli, but may not encode specific emotions (Adolphs, 2002; Schirmer & Kotz, 2006; Von Kriegstein & Giraud, 2004).

Collectively, these findings emphasize the crucial role of audiovisual integration in affect perception, as faces paired with a happy or angry voice exhibited distinctly different patterns of neural activity. Moreover, the location of this activity appears to closely correspond to those reported in the face mimicry literature (Adolphs, 2002; Bailey, Henry, & Varcin, 2012; Hess &

Fischer, 2014; Seidel, et al., 2010). In the current study, however, this effect was driven by the prosodic information, since all three FV conditions utilized the same set of face stimuli and differed only in the prosody of the voice they were paired with (angry, happy, neutral). These data suggest that while the physical qualities of a face play a major role in affect perception this is not entirely independent of the simultaneously presented vocal information. To further parse apart this relationship activity from the face only and prosody only conditions was compared to the angry and happy FV conditions. Interestingly, when compared to the face only condition, both the angry FV and happy FV conditions exhibited a similar distributed pattern of increased HbO activity over posterior-anterior areas in the right hemisphere (Figure 3.4a,b). Additionally, the FV conditions exhibited greater HbO levels than the prosody only conditions, but this activity was not as diffuse as that witnessed in the face comparisons (Figure 3.5a,b). Several additional comparisons were made, but their results were more difficult to interpret. These data are featured in figures 3.6, 3.7, and tables 3.5 and 3.6.

These findings highlight the inherent interrelatedness of affective vocal and facial expressions. What's more, these data support the notion that affect perception evolves in complexity, with initial processing originating in brain areas that are mutually responsive to all vocal (mSTG) and facial (ROR) displays of emotion. Subsequent integration with conceptually relevant knowledge appears to manifest as a divergence of neural activity, separating the right hemisphere into posterior and anterior subdivisions. These findings indicate that the processing of affective faces and voices may not be entirely separate.

The current study also has significant limitations that should be acknowledged. First, the involvement of subcortical structures in emotion perception cannot be assessed with NIRS due to limitations involving depth and spatial resolution (Hoshi, 2003; Ferrari & Quaresima, 2012).

Second, because we did not acquire electromyogram data the degree of facial mimicry/emotional contagion, if any, could not be assessed in our participants. Third, the lack of findings for the left dorsolateral frontal region should be taken with caution, as the technical failure of one source optode for some subjects necessitated the interpolation of 10 channels across the entire participant dataset. Finally, a technical error during the encoding of participants' responses made it impossible to analyze the corresponding behavioral data.

Despite these limitations, the data provided here provide an initial investigation of the neural correlates underlying emotion perception using a multimodal approach to increase the ecological validity and generalizability of the experiment to other neuroimaging studies. To our knowledge, this is the first fNIRS study to investigate multimodal emotion perception using a high-density optode array (Doi, Nishitani, & Shinohara, 2013; Bendall, Eachus, & Thompson, 2016). The striking anterior-posterior distinction between angry and happy FV conditions in the right hemisphere should be replicated and extended in future studies.

# REFERENCES


Adolphs, R., Damasio, H., Tranel, D., Damasio, A.R. (1996). Cortical systems for the recognition of emotion in facial expressions. *J Neuroscience 16*(23):7678-7687. PMID: 8922424

Adolphs, R. (2002). Recognizing emotion from facial expressions: psychological and neurological mechanisms. *Behavioral Cognitive Neuroscience* 1:21-61. PMID: 17715585

Adolphs, R., Tranel, D., Damasio, A.R. (2003). Dissociable neural systems for recognizing emotions. *Brain and Cognition 52*(1):61-69. PMID: 12812805

Bartlett, M.S., Viola, P.A, Sejnowski, T.J., Golomb, B.A., Larsen, J., Hager, J.C., Ekman, P. (1996). Classifying facial action. *Advances in Neural Information Processing Systems.* D. Touretzky, M. Mozer, & M. Hasselmo (Eds.), MIT Press, pp. 823-829.

Beauchamp, M.S., Argall, B.D., Bodurka, J., Duyn, J.H., Martin, A. (2004). Unraveling multisensory integration: patchy organization within human STS multisensory cortex. *Nature Neuroscience 7*(10):1190-1192. PMID: 15475952

Belin, P., Zatorre, R.J., Ahad, P. (2002). Human temporal-lobe response to vocal sounds. *Cognitive Brain Research* 13:17-26. PMID:11867247

Belin, P., Fillion-Bilodeau, S., Gosselin, F. (2008). The Montreal Affective Voices: a validated set of nonverbal affect bursts for research on auditory affective processing. *Behavioral Research Methods 40*(2):531-9. PMID: 18522064

Belin, P, Fecteau, S., Bédard, C. (2004). Thinking the voice: neural correlates of voice perception. *Trends in Cognitive Science 8*(3):129-135. PMID:15301753

Bendall, R.C., Eachus, P., Thompson, C. (2016). A Brief Review of Research Using Near-Infrared Spectroscopy to Measure Activation of the Prefrontal Cortex during Emotional Processing: The Importance of Experimental Design. *Frontiers Human Neuroscience* 10. PMID: 27812329

Borod, J.C., Pick, L.H., Hall, S., Sliwinski, M., Madigan, N., Obler, L.K., Welkowitz, J., Canino, E., Erhan, H.M., Goral, M., Morrison, C., Tabert, M. (2000). Relationships among, facial, prosodic, and lexical channels of emotional perceptual processing. *Cognition and Emotion 14*(2):193-211. DOI: 10.1080/026999300378932

Borod, J.C., Bloom, R.L., Brickman, A.M., Nakhutina, L., Curko, E.A. (2002). Emotional processing deficits in individuals with unilateral brain damage. *Appl. Neuropsychol.* 9, 23–36. PMID: 12173747

Bowers, D., Bauer, R.M., Coslett, H.B., Heilman, K.M. (1985). Processing of faces by patients with unilateral hemisphere lesions. *Brain Cogn 4*(3):258 –272. PMID: 4027060

Calvert, G.A., Bullmore, E.T., Brammer, M.J., Campbell, R., Williams, S.C., McGuire, P.K., Woodruff, P.W., Iversen, S.D., David, A.S. (1997). Activation of auditory cortex during silent lipreading. *Science 276*(5312):593-596. PMID:9110978

Capilla, A., Belin, P., Gross, J. (2013). The Early Spatio-Temporal Correlates and Task Independence of Cerebral Voice Processing Studied with MEG. *Cerebral Cortex* 23:1388-1395. doi:10.1093/cercor/bhs119

Davidson, R.J. (1995). Cerebral asymmetry, emotion, and affective style. In R.J. Davidson & K. Hughdahl (Eds.), *Brain Asymmetry* (pp. 361-387). Massachusetts: MIT Press.

de Gelder, B., Vroomen, J. (2000). The perception of emotions by ear and by eye. *Cognition and Emotion 14*(3):289-311. DOI: 10.1080/026999300378824

De Winter, F.L., Zhu, Q., Van den Stock, J., Nelissen, K., Peeters, R., de Gelder, B., Vanduffel, W., Vandenbulcke, M. (2015). Lateralization for dynamic facial expressions in human superior temporal sulcus. *Neuroimage* 106:340-352. PMID: 25463458

Decety, J., Lamm, C. (2007). The Role of the Right Temporoparietal Junction in Social Interaction: How Low-Level Computational Processes Contribute to Meta-Cognition. *The Neuroscientist 1*3(6):580-593. DOI: 10.1177/1073858407304654

Doi, D., Nishitani, S., Shinohara, K. (2013). NIRS as a tool for assaying emotional function in the prefrontal cortex. *Frontiers in Human Neuroscience* 7:770. PMID: 24302904

Ekman, P., Friesen, W.V. (1976). *Pictures of Facial Affect*. Palo Alto, CA: Consulting Psychological Press.

Ekman, P. (1992). An argument for basic emotion. *Cognition and Emotion 6*(3/4):169-200.

Ethofer, T., Van De Ville, D., Scherer, K., Vuilleumier, P. (2009). Decoding of emotional information in voice-sensitive cortices. *Current Biology* 19:1028–1033. PMID:19446457

Ethofer, T., Bretscher, J., Wiethoff, S., Bisch, J., Schlipf, S., Wildgruber, D., Kreifelts, B. (2013). Functional responses and structural connections of cortical areas for processing faces and voices in the superior temporal sulcus. *Neuroimage 1*(76):45-56. PMID:23507387

Fox, C.J., Young Moon, S., Iaria, G., Barton, J.J.S. (2009). The correlates of subjective perception of identity and expression in the face network: an fMRI adaptation study. *Neuroimage 44*(2):569-580. PMID:18852053

Früholz, S., Grandjean, D. (2013). Multiple subregions in superior temporal cortex are differentially sensitive to vocal expressions: A quantitative meta-analysis. *Neuroscience and Biobehavioral Reviews 37*(1):24-35. PMID: 23153796

Gandour, J., Tong, Y., Wong, D., Talavage, T., Dzemidzic, M., Xu, Y., Li, X., Lowe, M. (2004). Hemispheric roles in the perception of speech prosody. *Neuroimage* 23:344-357. PMID: 15325382

Gauthier, I., Tarr, M.J., Moylan, J., Skudlarski, P., Gore, J.C., Anderson, A.W. (2000). The Fusiform “Face Area" is Part of a Network that Processes Faces at the Individual Level. *J Cognitive Neuroscience 12*(3):495-504. PMID: 10931774

Gerdes, A.B.M., Wieser, M.J., Bublatzky, F., Kusay, A., Plichta, M.M., Alpers, G.W. (2013). Emotional sounds modulate early neural processing of emotional pictures. *Frontiers in Psychology 4*(741):1-12. PMID: 24151476

Gerdes, A.B.M., Wieser, M.J., Alpers, G.W. (2014). Emotional pictures and sounds: a review of multimodal interactions of emotion cues in multiple domains. *Frontiers in Psychology* 5. PMID: 25520679

Grandjean, D., Sander, D., Pourtois, G., Schwartz, S., Seghier, M.L., Scherer, K.R., Vuilleumier, P. (2005). The voices of wrath: brain responses to angry prosody in meaningless speech. *Nature Neuroscience 8*(2):145-156. PMID: 15665880

Haxby, J.V., Hoffman, E.A.,Gobbini, M.I. (2000). The distributed human neural system for face perception. *Trends Cognitive Science 4*(6):223-233. PMID:10827445

Hoffman, E.A., Haxby, J.V. (2000). Distinct representations of eye gaze and identity in the distributed human neural system for face perception**.** *Nature Neuroscience 3*(1):80-84. PMID: 10607399

Johnstone, T., van Reekum, C.M., Oakes, T.R., Davidson, R.J. (2006). The voice of emotion: an FMRI study of neural responses to angry and happy vocal expressions. *Scan* 1:242-249. doi:10.1093/scan/nsl027

Kanwisher, N., McDermott, J., Chun, M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J Neuroscience* 17: 4302±4311. PMID: 9151747

Kanwisher, N., Yovel, G. (2006). The fusiform face area: a cortical region specialized for the perception of faces. *Philosophical Transactions of the Royal Society B* 361:2109-2128. doi:10.1098/rstb.2006.1934

King. A.J., Nelken, I. (2009). Unraveling the principles of auditory cortical processing: can we learn from the visual system? *Nature Neuroscience 12*(6): 698–701. doi:10.1038/nn.2308

Kreifelts, B. Ethofer, T., Grodd, W., Erb, M., Wildgruber, D. (2007) Audiovisual integration of emotional signals in voice and face: an event-related fMRI study. *Neuroimage* 37:1445–1456. PMID: 17659885

Klasen, M., Chen, Y.H., Mathiak, K. (2012). Multisensory emotions: perception, combination and underlying neural processes. *Rev. Neurosci.* 23:381–392. doi: 10.1515/revneuro-2012-0040

Köchel, A., Plichta, M.M., Schäfer, A., Leutgeb, V., Scharmüller, W., Fallgatter, A.J., Schienle, A. (2011). Affective perception and imagery: A NIRS study. *Int J Psychophysiology 80*(3):192-197. PMID: 21419180

Korb, S., Frühholz, S., Grandjean, D. (2015). Reappraising the voices of wrath. *Social Cognitive and Affective Neuroscience 10*(12):1644-1660. PMID: 25964502

Kragel, P.A., LaBar, K.S. (2016). Decoding the Nature of Emotion in the Brain. *Trends in Cognitive Science 20*(6):444-455. PMID: 27133227

Kriegstein, K.,V., Giraud, A., (2004). Distinct functional substrates along the right superior temporal sulcus for the processing of voices. *Neuroimage* 22:948-955. PMID: 15193626

Leppänen, J.M., Nelson, C.A. (2009). Tuning the developing brain to social signals of emotions. *Nature Neuroscience 10*(1):37-47. PMID: 19050711

Liu, J., Harris, A., Mangini, M., Wald, L., Kwong, K., Kanwisher N. (2003). Distinct representations of faces in the FFA and the OFA: an fMRI study. In Society for Neuroscience. New Orleans, LA.

Massaro, D.W., Egan, P.B. (1996). Perceiving affect from the voice and the face. *Psychonomic Bulletin & Review 3*(2):215-221. PMID: 24213870

Patel, S., Scherer, K.R., Björkner, Sundberg, J. (2011). Mapping emotions into acoustic space: The role of voice production. *Biological Psychology 87*(1):93-98. PMID: 21354259

Peelen, M.V., Atkinson, A.P., Vuilleumier, P. (2010). Supramodal Representations of Perceived Emotions in the Human Brain. *J Neuroscience 30*(30):10127-10134. DOI:10.1523/JNEUROSCI.2161-10.2010

Pessoa, L., Adolphs, R. (2010). Emotion processing and the amygdala: from a 'low road' to 'many roads' of evaluating biological significance. *Nature Neuroscience 11*(11):773-782. PMID: 20959860

Pourtois, G., de Gelder, B., Vroomen, J., Rossion, B., Crommelinck, M. (2000). The time-course of intermodal binding between seeing and hearing affective information. *NeuroReport 11*(6): 1329-1333. PMID: 10817616

Rossion, B., Caldara, R., Seghier, M., Schuller, A. M., Lazeyras, F., Mayer, E. (2003). A network of occipito-temporal face-sensitive areas besides the right middle fusiform gyrus is necessary for normal face processing. *Brain 126*:2381–2395. PMID: 12876150

Rotshtein, P., Henson, R. N., Treves, A., Driver, J., Dolan, R. J. (2005). Morphing Marilyn into Maggie dissociates physical and identity face representations in the brain. *Nature Neuroscience* 8:107–113. doi:10.1038/nn1370

Samson, D., Apperly, I.A., Chiavarino, C., Humphreys, G.W. (2004). Left temporoparietal junction is necessary for representing someone else's belief. *Nature Neuroscience* 7:499-500. doi:10.1038/nn1223

Schirmer, A., Kotz, S.A. (2006). Beyond the right hemisphere: brain mechanisms mediating vocal emotional processing. *Trends in Cognitive Sciences 10*(1):24-30. PMID:16321562

Schirmer, A., Adolphs, R. (2017). Emotion Perception from Face, Voice, and Touch: Comparisons and Convergence *Trends in Cognitive Neuroscience 21*(3):216-228. PMID:28173998

Schröder, M. (2003). Experimental study of affect bursts. *Speech Communication* 40:99-116

von Kriegstein, K., Eger, E., Kleinschmidt, A., Giraud, A.L. (2003). Modulation of neural responses to speech by directing attention to voices or verbal content. *Brain Research Cognitive Brain Research 17*(1):48-55. PMID: 12763191

Vuilleumier, P., Pourtois, G. (2007). Distributed and interactive brain mechanisms during emotion face perception: Evidence from functional neuroimaging. *Neuropsychologia 45*(1):174-194. PMID: 16854439

Wildgruber, D., Hertrich, I., Riecker, A., Erb, M., Anders, S., Grodd, W., Ackermann, H. (2004). Distinct frontal regions subserve evaluation of linguistic and emotional aspects of speech intonation. *Cerebral Cortex 14*(12):1384-1389. PMID: 15217896

Wildgruber, D., Riecker, A., Hertrich, I., Erb, M., Grodd, W., Ethofer, T., Ackermann, H. (2005). Identification of emotional intonation evaluated by fMRI. *Neuroimage 24*(4):1233-1241. PMID:15670701

Wildgruber, D., Ackermann, H., Kreifelts, B., Ethofer, T. (2006). Cerebral processing of linguistic and emotional prosody: fMRI studies. *Progress in Brain Research* 156:249-268. DOI: 10.1016/S0079-6123(06)56013-

Winston, J. S., Vuilleumier, P., Dolan, R. J. (2003). Effects of low-spatial frequency components of fearful faces on fusiform cortex activity. *Curr. Biol.* 13:1824–1829. doi:10.1016/j.cub.2003.09.038

Wright, T.M., Pelphrey, K.A., Allison, T., McKeown, M.J., McCarthy, G., (2003). Polysensory interactions along lateral temporal regions evoked by audiovisual speech. *Cerebral Cortex* 13:1034–1043. PMID:12967920

Yang, D.Y.J, Rosenblau, G., Keifer, C., Pelphrey, K.A. (2015). An integrative neural model of social perception, action observation, and theory of mind. *Neuroscience & Biobehavioral Reviews* 51:263-275. PMID:25660957

Yovel, G., Kanwisher, N. (2004). Face perception: domain specific, not process specific. *Neuron* 44:747–748. doi:10. 1016/j.neuron.2004

# CHAPTER 4 – GENERAL DISCUSSION

Collectively, the present findings demonstrate that prosody is central to affect perception, with pronounced changes in behavioral responses and neural activity appearing to be linked to stimulus valence. The effect of stimulus modality was first investigated using a two-alternative forced-choice task, which showed that the happy face and voice (FV) condition and the face only condition exhibited higher percentages of happy responses when compared to the angry and neutral FV conditions, but were not significantly different from one another. These results indicate that while happy voices may bias responses to be 'happier' than the angry or neutral FV prosody conditions, the faces used in this study may be perceived to be inherently happier than they were intended to be. This finding questions the efficacy of the facial stimuli to equally represent both the happy and angry emotions, as well as the capacity of prosody to bias multimodal stimuli. Interestingly, the face only condition had the largest just noticeable difference (JND) value when compared to the prosody conditions, but boasted the fastest reaction times. These results seem counterintuitive as the JND has been used as a measure of confusion between choices, with smaller values indicating less confusion. These data may indicate that voices may provide supplementary information about the emotion of a stimulus which fosters more definitive responses. However, these responses appear to take more processing time.

The authors sought to further parse apart the relationship between affective faces and voices by using functional near-infrared spectroscopy (fNIRS) to study the neural correlates of the bimodal and unimodal influences on affect perception. These data showed that information appears to be uniquely combined in anterior and posterior areas of the right hemisphere, and these subdivisions are related to the valence of the presented stimulus. Moreover, these activations

persist even when compared to either the prosody or face stimuli shown alone, indicating that multimodal information is specially represented in the brain and this activity appears to be linked to the emotion of the simultaneously presented voice.

To further define this linkage, it is imperative to consider the findings from other neuroimaging modalities which have examined complementary measures of the same underlying neural activity. Evidence from electroencephalography (EEG) has provided keen insights into this endeavor, as EEG data is directly acquired from firing neurons, giving it excellent temporal resolution. Thus, EEG enables the close examination of multiple event-related potentials which may represent the progressive processing and integration of affective vocal and facial expressions. Multi-modal integration appears to take place over a longer, more variable timeframe than other components, starting approximately 110-200 ms after stimulus presentation (Pourtois, de Gelder, Vroomen, Rossion, & Crommelinck, 2000). During this period two positive components emerge, which respond to the emotional valence of a visual stimulus (P1, 100-130 ms) (Batty & Taylor, 2003; Calvo & Beltrán, 2013), and the prosody of a speaker's voice (P2, 110-250 ms) (Paulmann, Ott, & Kotz, 2011). A negative downward peak that's highly selective to images of faces emerges around 170 ms post-stimulus (N170) and is localized to the fusiform face area and occipital face area (Haxby, Hoffman, & Gobbini, 2000; Kanwisher, McDermott, & Chun 1997; Nguyen & Cunnington, 2013). Identifying these components is essential to understanding emotion perception as voices and faces appear to differentially effect brain activity, but it is unclear when this integration is initiated.

While the existing literature is somewhat limited, fNIRS has proven to be an effective tool in further dissociating the neural correlates of affect perception. Evidence from fNIRS suggests that emotion perception may develop early in life, as infants as young as 7-months of age exhibit

differences in both the time course and spatial distribution of oxygenated hemoglobin (HbO) over temporal areas when exposed to positively and negatively valenced faces (Nakato, Otsuka, Kanazawa, Yamaguchi, & Kakigi, 2011). Similarly, adults have been shown to exhibit increased HbO for unpleasant images, when compared to viewing pleasant images in bilateral prefrontal cortices (Hoshi, Huang, Kohri, Iguchi, Naya, Okamoto, & Ono, 2011). Additionally, Köchel and colleagues (2011), indicated that perception and imagery of visual affect may be differentially processed by parietal and occipital areas, respectively. While relatively fewer studies have focused on verbal affect perception, one study found that emotionally pleasant and unpleasant sounds exhibited greater HbO in the auditory cortex when compared to emotionally neutral sounds (Plichta, Gerdes, Alpers, Harnisch, Brill, Wieser, & Fallgatter, 2011). Together, these studies demonstrate the utility of fNIRS in affect research.

However, what should be emphasized is that much of NIRS research has been carried out using probes which record only portions of the cortex, most commonly: the prefrontal cortex (Hoshi, Huang, Kohri, Iguchi, Naya, Okamoto, & Ono, 2011; Doi, Nishitani, & Shinohara, 2013; Bendall, Eachus, & Thompson, 2016), temporal lobes (Plichta, et al., 2011), or parieto-occipital cortices (Köchel, Plichta, Schäfer, Leutgeb, Scharmüller, Fallgatter, & Schienle, 2011). The presented study sought to add to the present body of NIRS emotion research by implementing a high-density optode array to provide whole-head coverage, to capture the neural activity related to the visual, auditory, and audio-visual experience of emotion. This multimodal perspective was taken as relatively few studies have focused on verbal affect perception (Latinus & Belin, 2011; Schirmer & Adolphs, 2017) or its pairing with facial stimuli in fNIRS (Bendall, Eachus, & Thompson, 2016). This work is particularly relevant to populations known to exhibit impairments in social cognition (depression, autism spectrum disorders, schizophrenia). Deficits in recognizing,

identifying, and discriminating between emotions may be partially attributed to the abnormal integration of affective facial and vocal information.

The utility of the experiment and stimuli used in this study are threefold: 1) They are nonverbal which enables them to be used across multiple populations and ages, 2) Due to the hybrid block design, this experiment can be easily adapted for use in EEG, magnetoencephalogram, or functional magnetic resonance imaging research, and 3) The stimuli are freely available which encourages the push for reproduction and replication. The present study has presented good evidence that behavioral and hemodynamic measures of emotion perception can provide rich, compelling data about affect perception.

# REFERENCES


Batty, M., Taylor, M.J. (2003). Early processing of the six basic facial emotional expressions. *Brain Res Cogn Brain Res*. *17*(3):613-620. PMID: 14561449

Bendall, R.C., Eachus, P., Thompson, C. (2016). A Brief Review of Research Using Near-Infrared Spectroscopy to Measure Activation of the Prefrontal Cortex during Emotional Processing: The Importance of Experimental Design. *Frontiers Human Neuroscience* 10. PMID: 27812329

Calvo, M.G., Beltrán, D. (2013). Recognition advantage of happy faces: tracing the neurocognitive processes. *Neuropyschologia 51*(11):2051-61. PMID: 23880097

Doi, D., Nishitani S., Shinohara K. (2013). NIRS as a tool for assaying emotional function in the prefrontal cortex. *Frontiers in Human Neuroscience* 7:770. PMID: 24302904

Haxby, J.V., Hoffman, E.A., Gobbini, M.I. (2000). The distributed human neural system for face perception. *Trends Cognitive Science 4*(6):223-233. PMID:10827445

Hoshi, Y., Huang, J., Kohri, S., Iguchi, Y., Naya, M., Okamoto, T., et al. (2011). Recognition of human emotions from cerebral blood flow changes in the frontal region: a study with event-related near-infrared spectroscopy. *J Neuroimaging* 21, e94–e101 doi: 10.1111/j.1552-6569.2009.00454.x

Kanwisher, N., McDermott, J., Chun, M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J Neuroscience* 17:4302±4311. PMID: 9151747

Köchel, A., Plichta, M.M., Schäfer, A., Leutgeb, V., Scharmüller, W., Fallgatter, A.J., Schienle, A. (2011). Affective perception and imagery: A NIRS study. *Int J Psychophysiology 80*(3):192-197. PMID: 21419180

Latinus, M., Belin, P. (2011). Human voice perception. *Current Biology 21*(4):R143-145. PMID: 21334289

Nakato, E., Otsuka, Y., Kanazawa, S., Yamaguchi, M.K., Kakigi, R. (2011). Distinct differences in the pattern of hemodynamic response to happy and angry facial expressions in infants — A near-infrared spectroscopic study. *Neuroimage 54*(2):1600-6. PMID: 20850548

Nguyen, V.T., Cunnington, R. (2013). The superior temporal sulcus and the N170 during face processing: Single trial analysis of concurrent EEG–fMRI. *Neuroimage* 86:492-502. PMID: 24185024

Paulmann, S., Ott, D.V., Kotz, S.A. (2011). Emotional Speech Perception Unfolding in Time: The Role of the Basal Ganglia. *PLoS One 6*(3): e17694. PMID: 21437277

Plichta, M.M., Gerdes, A.B., Alpers, G.W., Harnisch, W., Brill, S., Wieser, M.J., Fallgatter, A.J. (2011). Auditory cortex activation is modulated by emotion: a functional near-infrared spectroscopy (fNIRS) study. *Neuroimage 55*(3):1200-7. PMID: 21236348

Pourtois, G., de Gelder, B., Vroomen, J., Rossion, B., Crommelinck, M. (2000). The time-course of intermodal binding between seeing and hearing affective information. *Neuroreport 11*(6):1329-33. PMID: 10817616.

Schirmer, A., Adolphs, R. (2017). Emotion Perception from Face, voice, and touch: Comparisons and convergence *Trends in Cognitive Neuroscience 21*(3):216-228. PMID:28173998