Running Head: COMMUNICATION OF EMOTION THROUGH HEAD MOTION

Communication of Emotional States through
Rigid Head Motion in Speakers and Singers

Sasha N. Ilnyckyj

McGill University

Abstract

Humans naturally perform body movements that facilitate communication during interpersonal vocal exchanges. This study investigated whether gross movements of the head can communicate the emotional state of an individual who is speaking or singing. Silent videos of vocalists speaking or singing emotional statements were manipulated to obscure the vocalists' facial features. These videos were presented to participants who were required to identify the emotional state (happy, sad or neutral) depicted in the video clip. Participants were capable of accurately identifying all three emotional categories above chance levels in both speech and song. Our findings suggest that rigid head motion may provide a visual cue that is sufficient for the recognition of emotional states during vocal communication.

Communication of Emotional States through Rigid Head Motion In Speech and Song

During conversation we naturally express ourselves through movement of our head, hands and body. This visual information can facilitate a variety of communicative functions. For instance, comprehension of speech is improved by the information available from lip movements - so called “lip-reading” (Jeffers & Barley, 1971). The intelligibility of speech produced in noise is also improved when a speaker’s face is visible (Summerfield, 1987). Furthermore, speech-associated arm and hand gestures have been found to improve both listener comprehension and speech production (Driskell & Radtke, 2003). A primary function of interpersonal communication is the portrayal of emotional states. Dynamic visual information provided by facial movements appears to enhance one’s ability to identify the emotions being expressed by speakers (Bassili, 1979; Kamachi et al., 2001; Cunningham & Wallraven, 2009). This work has focused almost exclusively on human facial movements. Yet, perceiving a speaker’s emotional state does not end with the face. Patterns of complex, extra-facial movement could provide a dynamic source of information that facilitates emotional perception. We seek to move away from ‘face-centric’ investigation and examine if movements made in other body areas, specifically the head, can be used to identify the emotional state of speakers. Can head motion convey emotional states?

Human emotional communication has been studied from a variety of perspectives. Investigation into the visual cues for emotion began with the human face (Darwin, 1872; Ekman, 1973). Individuals are remarkably accurate in recognizing emotions in facial expressions that are presented both in static images (Ekman, 1971) and in motion (Bieh et al., 1997; Massaro & Egan, 1996). Body movement has also been shown to signal an individual’s emotional state. For instance, Planalp (1996) had individuals record the specific cues they used to detect the emotional states of their house-mates. Nearly half of participants indicated that they used body language as a primary indicator of emotion. Montepare, Goldstein and Clausen (1987) found that subjects were capable of recognizing emotions such as sadness, anger and happiness based on a walker's gait. De Gelder (2004) recorded fMRI responses in individuals while they viewed full-body, still images of people in stances depicting fear, happiness and neutrality. Particularly when viewing fearful images, participants showed increased activity in areas associated with the processing of emotional information such as the amygdala, orbitofrontal cortex and anterior insula (for a review, see Adolphs, 2002). If gross body motions can reliably signal emotional states, localized head motions may provide similar cues.

In studies regarding verbal communication of emotion, research has focused on auditory features. For instance, Scherer (2003) has applied Brunswik's *functional lens model* of perception (Brunswik, 1956) to emotional communication in speech. This model emphasizes the importance of so-called *distal cues* - specific acoustic features such as speech rate, envelope, pitch height and intensity that serve as emotional signals for listeners. The model assumes that a speaker's emotional state is accompanied by physiological changes in respiration, phonation and articulation which produce emotion-specific modifications of speech features. Digital acoustic analysis (Banse & Scherer, 1996) and perceptual studies (E.g. Costanzo et al., 1969; Scherer & Oshinsky, 1977) have supported the notion that emotions have specific *auditory* profiles that characterize them. Similar work, however, has not yet been undertaken to identify dynamic *movement* profiles in the verbal communication of emotion.

The two forms of vocal communication of interest for this study are speech and song. There has been a multidisciplinary effort involving psychologists, linguists and engineers to investigate portrayal of emotion in *speech* (for a review, see Scherer, 2003). Yet, only recently has attention turned to emotional expression in *singing*. Thompson, Graham and Russo (2005) suggest that facial expressions made by singers during performance may reflect the emotional state of the performer and their reaction to musical features like dissonance. Livingstone, Palmer, Wanderley, Thompson and Lissemore (2011) found that participants were able to recognize emotions portrayed in song as accurately as those portrayed in spoken statements. Our objective is to observe the role played by the movements of speakers and singers in conveying emotion. Do the *movements* employed during song and speech differ? Are both equally effective at conveying emotion?

Some speech studies have documented links between head movements and the production of specific suprasegmental features (e.g. phrasing, stress, prosody). For instance, speakers' head motion can differentiate statements from questions (Fisher, 1969; Nicholson, Baum, Cuddy & Munhall, 2002) and determine word emphasis in phrases (Risberg & Lubker, 1978; Thompson, 1934). Munhall, Jones, Callan, Kuratate and Vatikiotis-Bateson (2004) used computer-generated, animated talking-heads to investigate the capacity for rigid head motion to convey linguistic information. In their study, the intelligibility of syllables spoken in noise was improved when natural, motion-captured head movements were applied to the animated figure compared to when motion was absent or distorted. Thus, movements of the head appear to provide useful supplementary information for perception of speech. Whether this extends to emotion or to song is still largely unknown.

Some music studies have also found links between performer's body movements and musical expression. Davidson (1993) employed *point-light display* figures (Johansson, 1973) to present recordings of violinists in three forms: lone audio, lone video or audio-video combined.

Participant ratings of expressivity were influenced substantially by the inclusion of visual input. In fact, *only* visual information was sufficiently nuanced to enable differentiation between normal and intentionally exaggerated performance manners. Vines, Krumhansl, Wanderley and Levitin(2006) presented either audio recordings or videos of clarinetists to participants who were asked to continuously rate the degree of musical tension they experienced. They found the visual modality conveyed information regarding musical tension to viewers that was independent from the tension viewers perceived when only the auditory modality was presented. They concluded that "musical-equivalents of *paralinguistic gestures* (such as head movements, eyebrow raising and postural adjustments) convey a significant amount of information that reinforces, anticipates or augments the auditory signal"(p. 107). This research demonstrates that body movements provide visual cues that may be equally important to auditory cues for communication of emotion, at least in musical performance.

The studies discussed above demonstrate the use of movement for expression by instrumentalists. Singers, unencumbered by the physical constraints of an instrument, may have an even wider array of expressive movements at their disposal. Singers' head motion has been linked to specific musical features. Thompson and Russo (2007) presented participants with silent video recordings of trained vocalists singing ascending melodic intervals. When participants were asked to rate on a numerical scale the size of the interval they believed was being sung, interval size ratings were positively correlated with the degree of head movement. Also, subsequent motion analysis showed a strong positive relationship (r =.94) between the size of the interval between the two notes and the magnitude of head displacement. Thus, they demonstrated a relationship between interval size and head movement both for the perception and production of sung phrases. An explanation is provided by Scherer (2003) who observed that increased vocal pitch range is associated with heightened arousal. Thus, singing a large pitch interval may connote heightened arousal that motivates head movement. If emotional singing is accompanied by increases in arousal this may manifest in head motion as well. Yehia, Kuratate and Vatikiotis-Bateson (2002) used experimentation and computational-modeling to investigate the relationship between fundamental frequency in speech and head movement. They found that a significant proportion of the variance (80-90%) in head motion could be accounted for by the frequency properties of spoken statements. Furthermore, they concluded that the coupling between head movements and fundamental frequency is not biomechanical (i.e. related to physically producing the sound). Rather, they believe the relationship to be *functional* such that the head movements somehow support communication. This supports the idea that extra-facial movements may play a facilitative role in verbal communication.

The work of Livingstone et al. (2009; 2011) has revealed interesting relationships between singers' motion and emotional states. In particular, specific facial movements seem to enhance the emotional intelligibility of utterances. Livingstone, Thomspon and Russo (2009) performed analysis on motion-captured movements of singers performing musical phrases with happy, sad or neutral emotional connotations. They found emotion-specific profiles in the movement of certain features. For instance, downward displacement of the lip corner occurred in performances intended to convey sadness. Research suggests that access to such visual cues can be important to perceiving emotions: viewers identified the emotions conveyed in speech and song more accurately when they watched videos than when they listened to audio recordings (Livingstone et al., 2011; Scherer, 2003). Livingstone et al. (2011) acknowledged the importance of broadening the scope of investigation to *extra*-facial features, particularly "[c]ues arising from rigid head motion”(p. 485), in an effort to identify new relationships between actions and communication of emotion.

The current study examined whether natural head motion that accompanies speech and song accurately convey emotional information. Participants attempted to identify, based on silent videos, emotions conveyed in short statements that were sung or spoken. The facial features of the vocalist were obscured with an opaque ellipsis in the videos, limiting visible movement to the surrounding head and neck. We hypothesized that participants would successfully identify the three emotional categories (happy, sad and neutral) above chance based solely on the head and neck movements they saw. Our expectation was that happiness and sadness would be identified with greater accuracy than neutrality. Furthermore, we hypothesized that accuracy would be enhanced for spoken vs. sung stimuli, based on the fact that head motion in singers appears to be influenced by musical factors such as interval size (Thompson & Russo, 2007) and fundamental frequency (Yehia et al., 2002), both of which are constrained in song but not in speech. If musical features that are unique to song drive head motion in a way that is unrelated to conveying specific emotions, singing may be detrimental to a clear communication of emotion. If this is so, observers should be less able to identify the emotion presented when a statement is being sung than when it is being spoken.

Methods

Participants

Twenty-five native English-speaking adults (18 female), ranging in age from 18-33 (*m*=21.63, SD=3.9) were recruited from the Montreal area. Participants had varying years of music instruction (*m*= 5.54, SD=5.34), singing training (*m*=1.71, SD=2.49) and acting experience (*m*=2.4, SD=2.07). No subjects had diagnosed vision or hearing problems. One participant was excluded from analysis as she recognized one of the vocalists in the stimulus videos. All subjects were recruited at McGill University through a classified posting. Ethical consent for the study was provided by the McGill University Research Ethics Board.

Stimuli

Stimuli consisted of silent video-recordings of vocalists whose faces had been occluded, speaking or singing short statements with various emotions. Four neutral English statements were used (“People going to the bank”, Children tapping to the beat”, “Children jumping for the ball”, “People talking by the door”). All were seven syllables in length and were matched in word frequency using the MRC psycholinguistic database (Coltheart, 1981).

Two vocalists (one female, one male) each with 10 years of musical experience acted as the model targets. Performers were recorded with a video camera (JVC Everio GZ-HD6 Camera) while speaking or singing each of the statements with three different emotional intentions: happy, neutral and sad. They performed two repetitions of each emotion. In the sung stimuli (*m* duration = 2.7 s), participants sang a fixed a 7-note isochronous melody with an I.O.I. of 300 ms (F4-F4-G4-G4-E4-E4-F4) based on a tempo provided by a metronome prior to recording. There were no pitch or duration constraints during the spoken stimuli (*m* = 1.8 s). Each participant recorded 48 statements (3 emotions x 2 productions(speech /song) x 2 repetitions x 4 statements) for a total of 96 stimuli.

The video recordings were then edited (Adobe Premiere After Effects & Adobe Premiere Elements - San Jose, CA) to occlude all the facial features. The occluding object covered the face and was a single opaque ellipsis, matched in size and skin-tone to each performer, as illustrated in Figure 1. Motion-tracking was used to ensure that the ellipsis followed the performer's movements and continued to cover the face throughout the video. The ellipsis spanned the entirety of the face including the jaw, chin and eyebrows. In the case of the male performer, the ellipsis was extended to occlude the laryngeal prominence ("Adam's apple") which may have provided information outside of rigid head motion alone. Videos included the performer's head, neck and upper chest while excluding the shoulders. As such, only head and neck movement was visible in each video. Each vocalist stood before a green background. The stimuli were exported as NTSC-format .AVI files (29 FPS at a resolution of 720x480 pixels) with no audio (muted).

Procedure & Design

Video-only (muted) stimuli with the faces occluded were presented to participants using E-Prime software (PST, Inc. – Sharpsburg, USA). The videos were viewed on a Dell U2410 wide-screen LCD monitor (Dell, Inc. - Round Rock, USA) as subjects sat comfortably at a distance of a few feet. Presentation of speech and song trails was blocked, with each block preceded by 12 practice trials (6 per vocalist). Experimental trials began with a short video clip preceded by a 500 ms. fixation cross in the center of the screen. Immediately following the video, participants identified the emotion expressed by the vocalist using a forced-choice bipolar rating scale (1=sad, 2=neutral, 3=happy). Participants also rated the strength of each emotion on a scale from 1(weakest)-7(strongest). Participants' responses were recorded on a standard PC keyboard. Between the first and second block an opportunity to rest was given. Following the trials, the participants completed a brief questionnaire on their musical and linguistic experience. Participation in the experiment took approximately 40 minutes, and participants received $10 in compensation.

The experiment was defined by a 2(Vocalist) x 2(Production: speech, song) x 3(Emotion: happy, neutral sad) x 2(Repetition) x 4(Statement) repeated-measures design with 96 trials per participant. Stimuli were blocked by Production (speech vs. song), with order counterbalanced such that half of participants began with a particular block. In each 48-video block the videos were presented in pseudorandom order to avoid presenting long runs of any particular condition. Analyses were collapsed across the random factors: Vocalist, Repetition and Statement.

Results

A two-way repeated measures analysis of variance (ANOVA) was performed to investigate the impact of *Emotion* (happy, sad & neutral) and *Production* (speech & song) on the participants' raw accuracy scores (coded in proportion correct, range=0-1). A significant main effect of *Emotion* (F(2, 46) = 26.494, p<0.0001) was observed, suggesting that accuracy differed between emotions (see Figure 2). Subsequent post-hoc tests (Tukey HSD=0.08, α=0.05) confirmed that accuracy for neutral videos (*m* accuracy = .80) was significantly higher than for happy videos (*m* = .70) which was, in turn, significantly higher than for sad videos (*m* = .56). Single-sample t-tests with Bonferroni correction ($t_{crit(46)}$=2.546, α=0.016) performed on each emotion confirmed that mean accuracy in all categories, sad (t=9.67), neutral (t=19.97) and happy (t=15.62), was significantly above chance levels (0.33).

A marginally significant main effect of *Production* was found (F(2,46) = 3.588, p = 0.071). Spoken utterances (*m*=0.70) were identified more accurately than sung utterances (*m* =0.66), providing partial support for our 2nd hypothesis. In addition, a significant interaction of *Emotion x Production* was observed (F(2, 46) = 6.567, p <0.005). As shown in Figure 3, the trend indicates the hypothesized pattern in happy and neutral emotions, in which spoken emotions were identified with greater accuracy than sung emotions.

We wished to confirm that participants' capacity to identify emotions was not simply a function of differences in the duration of the stimulus videos. The metronome that set the tempo during recording of the sung stimuli produced utterances with constrained duration (*m*=2.66 s, range=2.40-2.97 s) and little variance (SD=0.18 s). However, the spoken stimuli (*m*=1.81 s, range=1.30-2.90 s) had stimulus durations that were free to vary (SD=0.31 s), as illustrated in Figure 4. It is possible that the variable length of these spoken stimuli would influence the accuracy of viewers' emotion judgments; perhaps videos of greater duration provided more information, facilitating identification. To investigate this, a correlation was performed to test for a *positive* relationship between video duration and judgment accuracy. However, a negative correlation (*r* =-0.69, p<0.001) was found, indicating that viewers had improved accuracy for *shorter* stimuli. This addresses our concern that neutral spoken statements may have had impoverished information due to their reduced mean duration (*m*=1.58 s) compared to happy statements (*m*=1.82 s) and sad statements (*m*=2.02 s).

Discussion

Participants were able to identify all emotions conveyed in the videos well-above chance levels, even when access to facial expression and auditory information was eliminated. The results support our first hypothesis and indicate that rigid head motion may provide a visual cue sufficient for recognition of emotion during vocal communication. Our second hypothesis, that the accuracy of participants' emotional judgments would be superior in spoken utterances compared to sung utterances, was partially supported.

Our finding that head movement is capable of conveying specific emotions is consistent with prior research demonstrating the role of body motions in emotional communication (Planalp, 1987; Runeson & Frykholm, 1983). Rather than focussing on broad physical aspects like posture (de Gelder, 2004) or gait (Montepare, 1987), we limited the information provided to the movements of the head and neck that naturally accompany vocalizations. As individuals tend to focus most on the face and eyes during interpersonal communication (Kleinke, 1986), these adjacent regions may be particularly relevant to conveying emotional information. Furthermore, the ubiquity of spontaneous head motion during conversation (Hill & Johnston, 2001) suggests they serve a communicative function. Indeed, head movements have previously been associated with structural features of speech like word stress (Risberg & Lubker, 1978) and also appear to aid speech intelligibility (Munhall et al., 2004). Our results suggest that such movements also communicate emotional information, at least for broadly-defined emotions like happiness and sadness.

Participants were most accurate at identifying the emotion in neutral statements, followed by happy statements which were, in turn, better than sad statements. Admittedly, these results went against our *a priori* expectations that happiness would be conveyed best and neutrality worst. Our rationale was that happy utterances were generally accompanied by rapid, bobbing movements of the head that would be well-preserved even when facial features were occluded. We assumed that neutrality, conveyed generally through slow, small movements, would be frequently misidentified as sadness. Our results can be explained in reference to participants' comments made during the post-experiment debriefing. Neutrality was identified with the highest accuracy - approximately 80 percent - due to a characteristic *absence* of motion. Participants frequently expressed ease at identifying neutral stimuli because of their static nature, while admitting to finding happy-sad differentiation difficult. Happy and sad statements *both* contained movement and thus were more likely to be confused with one another. The negative correlation witnessed between the mean duration of videos and mean participant accuracy can be explained in reference to these findings. Neutral videos had the shortest mean duration and the *highest* accuracy, while sad videos had the longest mean duration and *lowest* accuracy. Both these relationships contributed to the negative trend of the correlation. This finding is informative to the extent in which it supports the notion that confusion of happy and sad movement cues affected participant accuracy. The greater duration of happy and sad videos represented a greater potential for participants to erroneously attribute certain motions to the wrong emotional category and produce identification errors. It is possible that after a certain point, more visual information may confuse judgment rather than aid it.

Importantly, *all* three emotional categories were identified at levels significantly higher than chance which addresses our primary question: head motion *can* convey emotional states. Subsequent investigation by the project collaborators will include motion analyses of the stimulus videos used. Perhaps, in a way similar to emotion-specific auditory profiles (Scherer, 2003) and facial feature displacement profiles (Livingstone et al., 2011), emotion-specific *head motion* profiles can be identified.

Head movements were capable of signalling emotion in both production modalities, speech and song. Based on evidence for coupling between head movements and musical features like intervals (Thompson & Russo, 2007) and pitch height (Yehia et al., 2003) we hypothesized that when vocalists sang the statements, their movements would be dictated in part by the melodic characteristics of the utterance. These constraints were expected to affect the vocalists' capacity to communicate emotion. Contrary to our second hypothesis, however, our analysis did not reveal a significant difference between the accuracy with which sung statements and spoken statements were identified. The comparison only approached significance. This suggests that even when singing, vocalists continued to make head movements that communicated their emotional state. Head movements come in various forms, including translational movements (i.e. changes in position) along multiple axes as well as rotational movements (i.e. changes in orientation) (Grossman, Leigh, Abel, Lanska and Thurston, 1998). The previous findings linking head motion with musical intervals and pitch height primarily looked at vertical translational displacements, measuring the magnitude of upward and downward motion accompanying singing of melodies. Emotion could be communicated by movements that are independent from vertical displacement such as 1) movement along a *different* axis 2) *rotational* movements or 3) certain *speed* and *timing* characteristics of movements. The planned motion analyses will be helpful for identifying the commonalities in movements between speech and song that enabled participants to recognize emotion in both with essentially equal accuracy. Our results suggest that head movement-based emotional communication is not domain-specific, but is utilized both when people speak and sing.

The current study has essentially been a successful test of principle. The results are encouraging and suggest pursuit of further research into emotional communication in rigid head movements may prove fruitful. Of course, all experimental designs can be improved. If we choose to expand this project, certain modifications would be prudent. Firstly, it would be beneficial to use more than two vocalists to produce the stimulus videos. Prior studies involving emotional perception have consistently found a high degree of inter-performer variability in emotional expressiveness. For instance, Davidson (1993) and Vines (2006) both found large differences in the extent to which their musical performers employed body motion while playing. These differences translated into large differences in the mean ‘expressiveness’ ratings assigned to different performers by both professional raters and experimental participants. To limit the impact of these variations on experimental results, it would be ideal to use a larger array of performers. Furthermore, a larger group of vocalists would limit the number of experimental trials contributed by each and reduce potential learning effects in which participants ‘become wise’ to how an individual communicates a certain emotion.

Another change worth considering would be to increase the number of emotions presented. ‘Happy’ and ‘sad’ are broad descriptors that many would argue are more akin to emotional ‘families’ (Ekman, 1992a) than singular emotions. Ekman (1992a) describes these families as groups of related affective states with similarities in the expression, physiological activity and situational antecedents that underlie them. For instance, the ‘anger’ family would include states such as annoyance, resentment, indignation and rage. As such, it would be interesting to include a number of more nuanced emotional states into the design to see if, for instance, participants could tell contentment apart from elation or grief from disappointment. An added benefit of this approach would be an increase in the number of potential categories participants could select. In this case, the task would be less about *discrimination* (i.e. “this is *not* happy and *not* neutral, therefore it must be sad”) and more a test of *recognition*.

Finally, a potential off-shoot of this current study would be to investigate the degree to which head movements affect emotional identification when auditory information *is* present. Our present study used silent videos, but by including audio information it would be possible to investigate the relative influence of visual and auditory information on perception of emotional states during speech and song. Prior research in audio-visual integration has documented perceptual phenomena such as the so-called “fusion illusion” demonstrated in the McGurk effect (McGurk & MacDonald, 1976). The McGurk effect occurs when mismatched visual and acoustic events are perceptually integrated. Viewers are shown a manipulated audio-video recording of individuals speaking single syllables. When the auditory channel provides the syllable /ba/ and the visual channel presents a speaker mouthing /ga/, viewers consistently report hearing a syllable that is an intermediate between the two: /da/. This occurs even when the acoustic signal is perfectly clear, which suggests that incongruent visual information is capable of influencing perception in other modalities. Testing for a similar effect in head-based emotional communication would involve presenting occluded stimulus videos similar to those in this experiment with either corresponding or mismatched audio recordings. For example, if sad head motion were to be paired with happy audio, how might this affect participants' judgments? Would they favour vision or audition? Would their impression of emotional strength and intensity be modified? This, and other cross-modal perception paradigms, may be a worth-while direction to pursue.

In sum, these findings indicate that rigid head motion may be a means by which individuals communicate their emotional state during interpersonal vocal communication. Furthermore, this cue does not seem to be domain-specific, at least between speech and song, as head movements reliably signal emotional states in both modalities investigated. Participants demonstrated high task competency within our relatively-brief experiment which suggests that identifying emotion based on head-movement is somewhat intuitive and requires no specific training. In daily interpersonal communication, we are often in situations where visual and auditory information is incomplete (e.g. when conversing in large groups, when facial features are occluded by sunglasses, masks, scarves, etc. or when in noisy or crowded environments). Perhaps in such situations, rigid head-movements are employed as visual signals that aid perception and identification of emotional states. Investigation of these dynamic cues stands to contribute a novel perspective to the literature on the communicative function of human body movement.

Acknowledgements

The author wishes to thank Dr. Caroline Palmer and Dr. Steven R. Livingstone for their guidance, patience and expertise throughout this process. Additional thanks go to Jennifer Lissemore for her technical prowess and to the Canada Research Chair grant for supporting this research.

References

Adoplhs, R. (2002). Recognizing Emotion from Facial Expressions: Psychological and Neurological Mechanisms. *Behavioural and Cognitive Neurosci. Reviews, 1(1),* 21-62.

Banse, R., & Scherer, K.R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology, 70(3),* 614-636.

Bassili, J.N. (1979). Emotion Recognition: The Role of Facial Movement and the Relative Importance of Upper and Lower Areas of the Face. *Journal of Personality and Social Psychology, 37(11),* 2049-2058.

Biehl, M., Matsumoto, D., Ekman, P., Hearn, V., Heider, K., & Kudoh, T.(1997). Matsumoto andEkman's Japanese and Caucasian Facial Expressions of Emotion (JACFEE): Reliability Data and Cross-National Differences. *Journal of Nonverbal Behavior, 21,* 3-21.

Brunswik, E. (1956). *Perception and the Representative Design of Psychological Experiments*. Berkeley, CA: University of California Press.

Coltheart, M. (1981). The MRC Psycholinguistic Database. *Quarterly Journal of Experimental Psychology, 33A,* 497-505.

Costanzo, F.S., Markel N.N., & Costanzo, P.R. (1969). Voice quality profile and perceived emotion. *Journal of Counselling Psychology, 16,* 267-270.

Cunningham, D.W., & Wallraven, C. (2009). Dynamic information for the recognition of conversational expressions. *Journal of Vision, 9(13):7,* 1-17.

Darwin, C. (1872). *The Expressions of Emotion in Man and Animals*. John Murray, London (third ed., P. Ekman (Ed.), 1998, Harper Collins, London)

Davidson, J. (1993) Visual perception of performance manner in the movements of solo musicians.*Psychology of Music, 21,* 103-113.

de Gelder, B., Snyder, J., Greve, D., Gerard, G., & Hadjkhani, N. (2004). Fear fosters flight: A mechanism for fear contagion when perceiving emotion expressed by a whole body. *Proceedings of the National Academy of Sciences, USA, 101,* 16701-16706.

de Gelder, B. (2009). Why bodies? Twelve reasons for including bodily expressions in affective neuroscience. *Phil. Trans. R. Soc. B, 364,* 3475-3484.

Driskell, J.E, & Radtke, P.H. (2003). The Effect of Gesture on Speech Production and Comprehension. *Human Factors: The Journal of the Human Factors and Ergonomics Society,* 45(3), 445-454.

Ekman, P., & Friesen, W.V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology, 17,* 124-129.

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

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

Grossman, G.E., Leigh, R.J., Abel, L.A., Lanska, D.J., & Thurston, S.E. (1998). Frequency and velocity of head perturbations during locomotion. *Experimental Brain Research, 70,* 470-476.

Fisher, C.G. (1969). The visibility of terminal pitch contour. *Journal of Speech and Hearing Research, 12,* 379-382.

Hill, H. & Johnston, A. (2001). Categorizing sex and identity from the biological motion of faces. *Current Biology, 11(11),* 880-885.

Jeffers, J., & Barley, M. (1971). *Speechreading (Lip-reading*). Springfield, IL: Charles, C. Thomas.

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception and Psychophysics, 14,* 201-211.

Kamachi, M., Bruce, V., Mukaida, S., Gyoba, J., Yoshikawa, S., & Akamatsu, S. (2001). Dynamic properties influence the perception of facial expressions. *Perception, 30,* 875-887.

Kleinke, L. (1986). Gaze and eye contact. *Psychological Bulletin, 100,* 78-100.

Livingstone, S. R., Thompson, W.F., & Russo, F. (2009). Facial Expressions and Emotional Singing: A Study of Perception and Production with Motion Capture and Electromyography. *Music Perception, 26(5),* 475-488.

Livingstone, S.R., Palmer C., Wanderley M.M., Thompson, W.F., & Lissemore, J. (2011). Facial expressions in vocal performance: Visual communication of emotion. *Proceedings of the International Symposium on Performance Science, USA,* 545-550.

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

Massaro D.W., & Egan, P.B. (1996). Perceiving Affect from the Voice and the Face. *Psychonomic Bulletin and Review, 3(2),* 215-221.

Montepare, J.M., Goldstein, S.B., & Clausen, A. (1987). The Identification of Emotions from Gait Information. *Journal of Nonverbal Behaviour, 11(1),* 33-42.

Munhall, K.G., Jones, J.A., Callan, D.E., Kuratate, T., & Vatikiotis-Bateson, E. (2004). Visual Prosody and Speech Intelligibility: Head Movement Improves Auditory Speech Perception. *Psychologial Science, 15(2)*, 133-137.

Nicholson, K.G., Baum, S., Cuddy L., & Munhall, K.G. (2002). A case of impaired auditory and visual speech prosody perception after right hemisphere damage. *Neurocase, 8,* 314-322.

Planalp, S. (1996). Varieties of Cues to Emotion in Naturally Occurring Situations. *Cognition and Emotion*, 10(2), 137-154.

Risberg, A., & Lubker, J. (1978). Prosody and speechreading. *Speech Transmission Laboratory Quarterly Progress Report and Status Report, 4,* 1-16.

Runeson, S., & Frykholm, G. (1981). Visual perception of lighted weight. *Journal of Experimental Psychology Human Perception and Performance, 7,* 733-740.

Scherer, K.R., & Oshinsky, J.S., 1977. Cue utilization in emotion attribution from auditory stimuli. *Motiv. Emot., 1,* 331-346.

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

Summerfield, A.Q. (1987). Some preliminaries to a comprehensive account of audio-visual speech perception. In B. Dodd & R. Campbell (Eds.), *Hearing by eye* (pp. 3-51). London: Erlbaum Associates.

Thompson, D.M. (1934). On the detection of emphasis in spoken sentences by means of visual, tactual and visual-tactual cues. *Journal of General Psychology, 11,* 160-172.

Thompson, W.F., Graham P., & Russo, F.A. (2005). Seeing music performance: Visual influences on perception and experience. *Semiotica*, *156,* 203-227.

Thompson, W.F., & Russo, F. (2007). Facing the Music. *Psychological Science, 18(9),* 756-757.

Vines, B.W., Krumhansl, C.L., Wanderley, M.M., & Levitin, D.J. (2006). Cross-modal Interactions in the perception of musical performance. *Cognition, 101,* 80-113.

Yehia, H., Kuratate, T., & Vatikiotis-Bateson, E. (2002). Linking facial animation, head motion and speech acoustics. *Journal of Phonetics, 30,* 555-568.

Figures

Figure 1. Sample trial produced by female vocalist.

Figure 2. Mean proportion of correct judgments as a function of emotion presented.

Figure 3. Mean proportion of correct judgments as a function of emotion presented and production mode (speech/song).

Figure 4. Mean video duration as a function of emotion presented and production mode (speech/song).

Figure 1

Figure 2

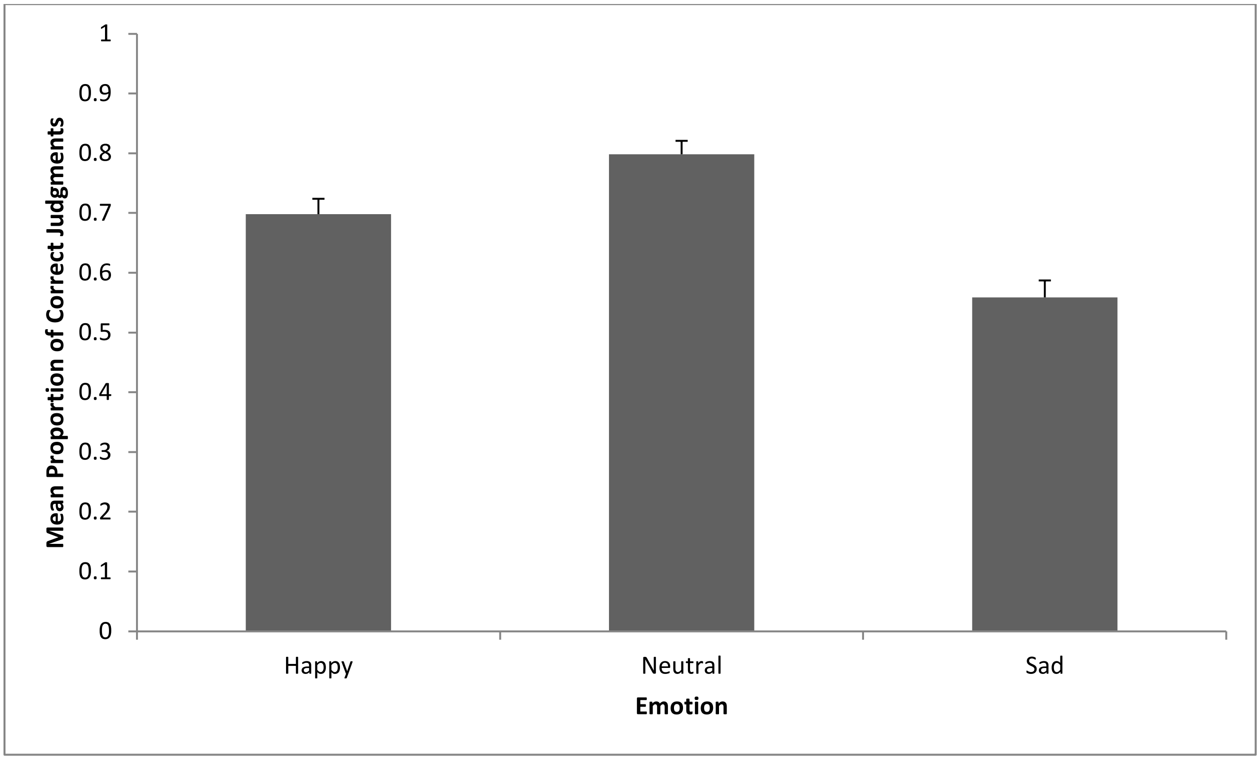

1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Mean Proportion of Correct Judgments
Happy
Neutral
Sad
Emotion

Figure 3

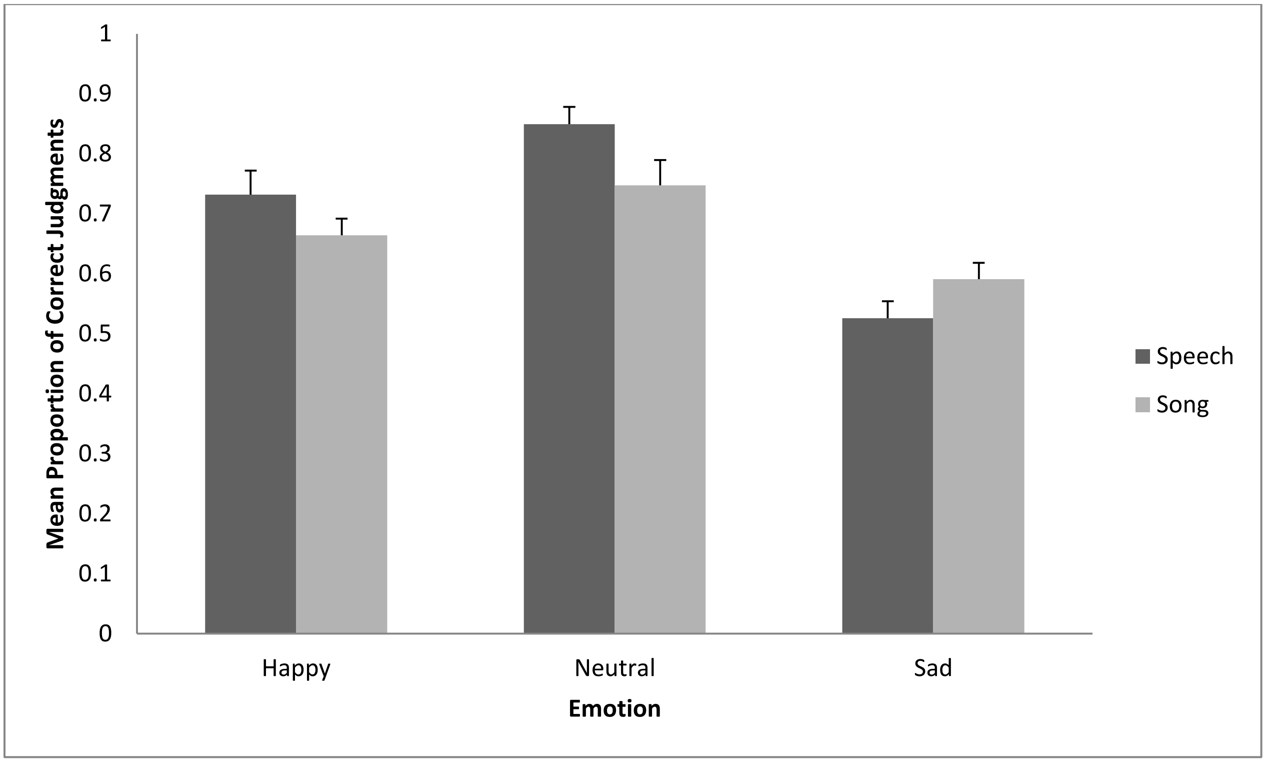
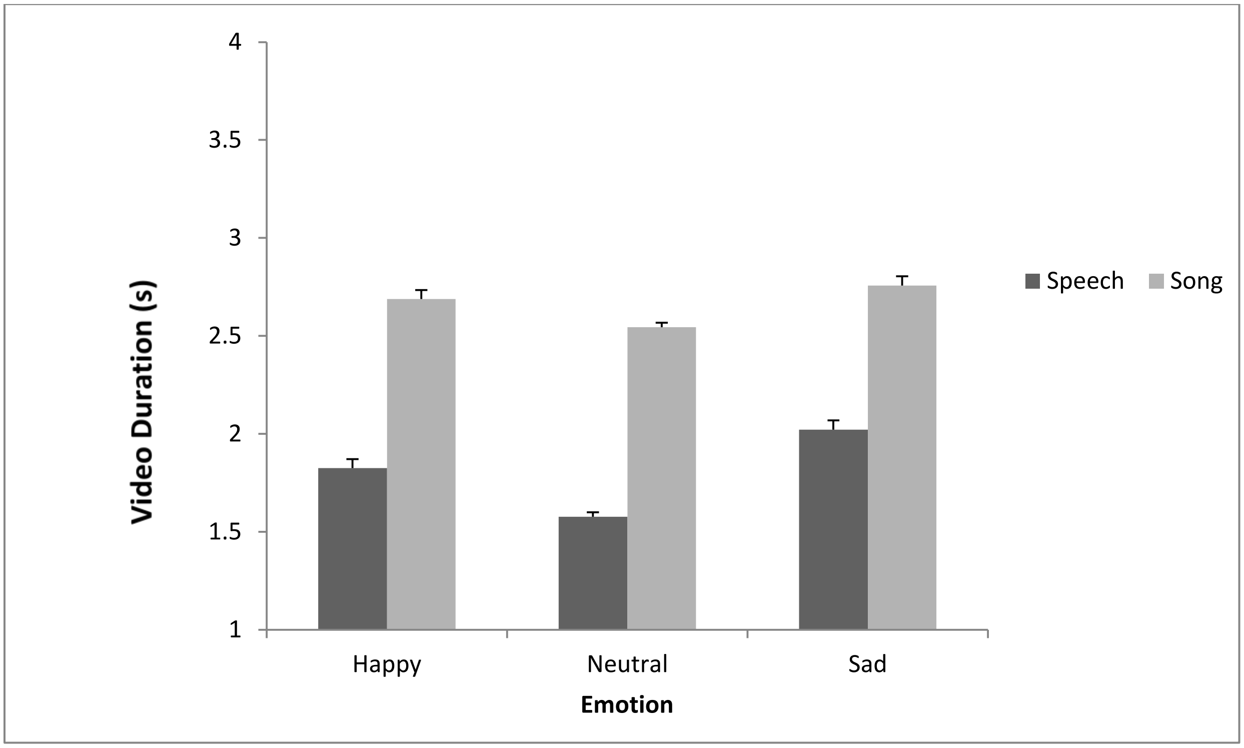

1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Mean Proportion of Correct Judgments
Happy
Neutral
Sad
Emotion
■ Speech
■ Song

Figure 4


4
3.5
3
2.5
2
1.5
1
Video Duration (s)
Happy
Neutral
Sad
Emotion
■ Speech
■ Song