Facial Expressions and Emotional Singing: A Study of Perception and Production with Motion Capture and Electromyography

Steven R. Livingstone
McGill University

William Forde Thompson
Macquarie University

Frank A. Russo
Ryerson University

digital.library.ryerson.ca/object/138

Please Cite:
Livingstone, S. R., Thompson, W. F., & Russo, F. A. (2009). Facial expressions and emotional singing: A study of perception and production with motion capture and electromyography. *Music Perception: An Interdisciplinary Journal, 26*(5), 475–488.
doi:10.1525/mp.2009.26.5.475

Ryerson University Library

library.ryerson.ca

Facial Expressions and Emotional Singing: A Study of Perception and Production with Motion Capture and Electromyography

Steven R. Livingstone
McGill University, Montreal, Canada

William Forde Thompson
Macquarie University, Sydney, Australia

Frank A. Russo
Ryerson University, Toronto, Canada

FACIAL EXPRESSIONS ARE USED IN MUSIC PERFORMANCE to communicate structural and emotional intentions. Exposure to emotional facial expressions also may lead to subtle facial movements that mirror those expressions. Seven participants were recorded with motion capture as they watched and imitated phrases of emotional singing. Four different participants were recorded using facial electromyography (EMG) while performing the same task. Participants saw and heard recordings of musical phrases sung with happy, sad, and neutral emotional connotations. They then imitated the target stimulus, paying close attention to the emotion expressed. Facial expressions were monitored during four epochs: (a) during the target; (b) prior to their imitation; (c) during their imitation; and (d) after their imitation. Expressive activity was observed in all epochs, implicating a role of facial expressions in the *perception, planning, production,* and *post-production* of emotional singing.

Received September 8, 2008, accepted February 14, 2009.

Key words: music cognition, emotion, singing, facial expression, synchronization

There is behavioral, physiological, and neurological evidence that music is an effective means of communicating and eliciting emotion (Blood & Zatorre, 2001; Juslin & Sloboda, 2001; Juslin & Västfjäll, 2008; Khalfa, Peretz, Blondin, & Manon, 2002; Koelsch, 2005; Rickard 2004). The emotional power of music is reflected in both the sonic dimension of music as well as the gestures and facial expressions used by music performers (Thompson, Graham, & Russo, 2005; Thompson, Russo, & Quinto, 2008). There is a vast body of research on the acoustic attributes of music and their association with emotional communication (Juslin, 2001). Far less is known about the role of gestures and facial expressions in the communication of emotion during music performance (Schutz, 2008).

This study focused on the nature and significance of facial expressions during the *perception, planning, production,* and *post-production* of emotional singing. When individuals perceive an emotional performance, perceptual-motor mechanisms may activate a process of mimicry, or synchronization, that involves subtle mirroring of observable motor activity (Darwin, 1872/1965; Godøy, Haga, & Jensenius, 2006; Leman, 2007; Molnar-Szakacs & Overy, 2006). When musicians are about to sing an emotional passage, advanced *planning* of body and facial movements may facilitate accurate performance and optimize expressive communication. When musicians engage in the *production* of an emotional performance, facial expressions support or clarify the emotional connotations of the music. When musicians complete an emotional passage, the bodily movements and facial expressions that were used during production may linger in a *post-production* phase, allowing expressive communication to persist beyond the acoustic signal, and thereby giving greater impact and weight to the music.

Perceivers spontaneously mimic facial expressions (Bush, Barr, McHugo, & Lanzetta, 1989; Dimberg, 1982; Dimberg & Lundquist, 1988; Hess & Blairy, 2001; Wallbott, 1991), even when facial stimuli are presented subliminally (Dimberg, Thunberg, & Elmehed, 2000). They also tend to mimic tone of voice and pronunciation (Goldinger, 1998; Neumann & Strack, 2000), gestures and body posture (Chartrand & Bargh, 1999), and breathing rates (McFarland, 2001; Paccalin & Jeannerod, 2000). When an individual perceives a music performance, this process of facial mimicry may function to facilitate rapid and accurate decoding of music structure and emotional information by highlighting relevant visual and kinaesthetic cues (Stel & van Knippenberg, 2008).

Music Perception volume 26, issue 5, pp. 475–488, issn 0730-7829, electronic issn 1533-8312 © 2009 by the regents of the university of california. all rights reserved. please direct all requests for permission to photocopy or reproduce article content through the university of california press's rights and permissions website, http://www.ucpressjournals.com/reprintinfo.asp. DOI:10.1525/mp.2009.26.5.475

It is well established that attending to facial features facilitates speech perception, as illustrated in the ventriloquism effect (Radeau & Bertelson, 1974), and the McGurk effect (McGurk & MacDonald, 1974). In the McGurk effect, exposure to a video with the audio syllable "ba" dubbed onto a visual "ga" leads to the perceptual illusion of hearing "da." The findings implicate a process in which cues arising from different modalities are perceptually integrated to generate a unified experience. The effect may occur automatically and preattentively: it is observed regardless of whether observers attempt to selectively attend to auditory information or whether they are distracted by a secondary task (de Gelder, Vroomen, & Pourtois, 1999). Audio-visual integration plays a significant role in maintaining verbal decoding in the face of decreasing signal-to-noise ratio.

The significance of audio-visual integration extends beyond speech perception. Accompanying visual information affects judgments of a wide range of sound attributes, including the loudness of clapping (Rosenblum, 1988), the onset characteristics of tones (Saldaña & Rosenbaum, 1993), the depth of vibrato (Gillespie, 1997), tone duration (Schutz & Lipscomb, 2007), and the perceived size of melodic intervals (Thompson & Russo, 2007). Visual aspects of music also affect perceptions of the emotional content of music. In an early investigation, Davidson (1993) asked participants to rate the expressiveness of solo violinist performances. Musicians performed in three ways: deadpan, projected (normal), and exaggerated. Participants rated the expressivity of point-light displays of performances under three conditions: audio only, visual only, and audio-visual. Expressive intentions were more clearly communicated by visual information than by auditory information.

Facial expressions play a particularly important role in the communication of emotion (Thompson et al., 2005; Thompson et al., 2008). Recently, Thompson et al. (2008) presented audio-visual recordings of sung phrases to viewers, who judged the emotional connotation of the music in either single-task or dual-task conditions. Comparison of judgments in single- and dual-task conditions allowed an assessment of the role of attention in audio-visual integration. Results suggested that perceivers preattentively integrate emotional cues arising from the acoustic dimension of music with cues arising from the facial expressions of singers, giving rise to an interpretation that reflects a balance of auditory and visual cues. Other investigations involving instrumental and vocal music have yielded similar results (Dahl & Friberg, 2004; 2007; Davidson & Correia, 2002; Geringer, Cassidy & Byo, 1997; for a review, see Shutz, 2008).

Facial mimicry influences the recognition of emotional facial expressions (Niedenthal, Brauer, Halberstadt, & Innes-Ker, 2001; Stel & van Knippenberg, 2008). For example, holding the face motionless reduces the experience of emotional empathy (Stel, van Baaren, & Vonk, 2008). Recognition of emotional facial expressions is also slower for individuals who consciously avoid facial movements than for those who avoid other types of movements, but are free to make facial movements (Stel & van Knippenberg, 2008; see also Pietromonaco, Zajonc, & Bargh, 1981; Wallbott, 1991).

The Mirror Neuron System (MNS) has been proposed as a neurological candidate for facial mimicry (Di Pellegrino, Fadiga, Fogassi, Gallese, & Rizzolatti, 1992; Ferrari, Gallese, Rizzolatti, & Fogassi, 2003; Gallese, Fadiga, Fogassi, & Rizzolatti, 1996; Kohler et al., 2002; Lahav, Saltzman, & Schlaug, 2007; for a review, see Rizzolatti & Craighero, 2004). The MNS may also underpin a Theory of Mind (ToM): the ability to understand the mental and emotional state of conspecifics by stepping into their "mental shoes" (Humphrey, 1976; Premack & Woodruff, 1978). One model of ToM is *Simulation Theory*. According to Simulation Theory, the MNS functions to produce a simulation of a conspecific's mental and emotional state (Gallese, 2001; Gallese & Goldman, 1998; Gordon, 1986; Heal, 1986; Oberman & Ramachandran, 2007; Preston & de Waal, 2002). Livingstone and Thompson (2009) proposed that Simulation Theory and the MNS can account for our emotional capacity for music, and its evolutionary origins. Molnar-Szakacs and Overy (2006) also highlight the role of the MNS, suggesting that it underlies our capacity to comprehend audio, visual, linguistic, and musical signals in terms of the motor actions and intentions behind them.

As music is temporally structured, understanding the time course of emotional communication in music is particularly important (Bigand, Filipic, & Lalitte, 2005; Schubert, 2001; Vines, Krumhansl, Wanderley, & Levitin, 2006; Wanderley, 2002; Wanderley, Vines, Middleton, McKay, & Hatch, 2005). To our knowledge, no studies to date have analyzed the time course of emotional facial expressions during vocal performance. Such investigations may shed light on motor planning and memory behaviors that underlie the communication of emotion in music (Palmer, 2005, 2006). Based on recent research and theory, we predict that emotional facial expressions will be evident not only during the actual production of emotional song, but also in the perception, planning, and post-production of emotional song.

In two experiments, we analysed emotional facial expressions in vocal music performance. Both experiments adopted the *emotional-production synchronization paradigm*, developed by Thompson, Livingstone, and Palmer (2008). In this technique, skilled singers are exposed to a videorecording of another musician singing a short musical phrase with varying emotional intentions. The participant then imitates the sung phrase, paying close attention to the emotional connotations of the target performance. Facial expressions may then be analyzed at several time periods or epochs: during the perception of the target, as the participant prepares to imitate the target performance, during the imitation itself, and immediately following the imitation. In our investigation, facial expressions were monitored using motion capture (Experiment 1) and EMG (Experiment 2). Motion capture allowed an analysis of the facial movements associated with emotional singing, whereas EMG allowed an analysis of the muscular activity that underlies emotional singing (Hess & Blairy, 2001; Sonnby-Borgstrom, Jonsoon, & Svensson, 2003).

Experiment 1

Method

PARTICIPANTS

Eight musically trained university students from Macquarie University participated in this study. Seven were students in the music department and acted as participants (4 men and 3 women, $M = 23.29$ years, $SD = 2.56$, range = 20-27 years). All participants had music experience ($M = 10.00$ years, $SD = 8.93$, range 1-25 years), and most had some formal music training ($M = 4.29$ years, $SD = 4.57$, range 0-13 years). Four participants reported that they had theatre experience. The eighth participant was from the psychology department and acted as the model participant. The model participant (50 year old woman) had 8 years of formal singing lessons and 40 years of music experience. Testing took approximately 120 minutes, with participants paid $40 AUD for their participation.

STIMULI AND MATERIALS

Stimuli were constructed from 18 video recordings of the model participant singing six sentences, each performed with three emotional intentions: happy, sad, and neutral. The verbal content of the six phrases was identical for each of the three emotion conditions (e.g., grass is green in summertime).

Each melodic phrase was sung in 4:4 time at 78 bpm. The same melody was used for all sentences and consisted of one bar of music with the following pitches at each successive beat: tonic, supertonic, leading note, and tonic (F4, G4, E4, F4). Each phrase consisted of seven syllables, with a different syllable sung on successive eighth notes and the final syllable sung on beat 4 (e.g., beat 1 = grass is; beat 2 = green in; beat 3 = summer; beat 4 = time). Sung phrases were recorded using a Sony Handycam DCR-TRV19E. A light and audio emitting metronome, set to 78 bpm, was attached to the camera to facilitate accurate timing by the singer.

Training stimuli were constructed from 18 video recordings of the model participant. The training stimuli were produced in a manner that is consistent with the experimental stimuli but with a slightly modified melody (F4, G4, C4, F4) to reduce a practice effect. Each recorded song was edited using Final Cut Pro 6 and inserted into the presentation stimuli. Each presentation stimulus defined 4 epochs, each lasting one 4:4 bar. Prior to epoch 1, a black screen was displayed for 1.5 seconds, with a 0.2 s warning beep occurring after 1.3 s. Epoch 1 was the recording of the model participant (target song) and faded to black during the last 0.5 s. Epoch 2 was a count-down bar that allowed participants time to prepare for their imitation. It displayed the numbers 4-3-2-1 in the centre of the screen, with each number synchronised to an audible beep from a metronome. Epoch 3 was the imitation bar and displayed a green dot for 0.5 s at the start of the bar to signal the commencement of the imitation. Epoch 4 consisted of four beats of silence with a black screen, followed by a 0.2 s audible beep that signaled the end of the trial. Each presentation lasted 13 s.

A three dimensional passive optical motion capture system (Vicon MX+ with 4 MX-F20 2-megapixel infrared cameras, analog-to-digital external patch panel, and MX Ultranet HD) captured the motion of participants' facial expressions and vocal utterances, operating at a frame rate of 200 Hz. Thirteen reflective markers were placed on the participant's face, consisting of three 9 mm diameter spherical markers, and ten 4 mm diameter hemispherical markers. Markers were placed on the forehead, inner and middle of each eyebrow, nose-bridge, nose tip, left and right sides of the head, upper and lower lip, and left and right lip corners.

Target stimuli were viewed on a 15'' MacBook Pro, running OS X 10.5.4 (Leopard), with external Altec Lansing AVS200B speakers. Stimuli were output using Leopard's in-built *Quick Look* function. The output from the speakers (target stimuli) and the vocal utterances of the participants were captured using a Rode K2 condenser microphone. Signal gain was applied using an MBOX2 audio interface, and was output to a

Vicon analog input device for synchronous digitization with motion capture data.

PROCEDURE

Participants were seated in front of the four-camera array, with the viewing laptop elevated to eye height. They were instructed to observe the target song on each trial, paying close attention to the emotion being expressed, visually and aurally. They were told to prepare for their imitation in bar 2 (the count-down epoch) and then imitate the sung melody and expressed emotion as effectively as possible from the first beat of bar 3 (imitation epoch). Participants could express the target emotion in their own way, however. That is, they were not required to mimic individual features of facial expressions. At the end of each trial, they selected from a list of terms the emotion expressed in the target stimulus (happy, angry, fear, sadness, or neutral). This procedure was done to ensure that participants recognized and reproduced the intended emotion. Participants achieved identification accuracy of 97.62%.

Prior to the experiment, a training session consisting of 18 trials was conducted. Presentation of training stimuli was randomized. Experimental stimuli, also randomized, were then presented to the participant. An additional set of 18 captures was then conducted with the experimental stimuli, to ensure that a clean capture was obtained for each stimulus instance.

MOVEMENT ANALYSIS

Raw capture data were reconstructed using Vicon Nexus 1.3.109. Occasional missing data in facial marker movement were interpolated with spline curve fitting. Digitized voltage data from the microphone input were used to synchronize each capture with the viewed stimulus, with start and end beeps used as landmarks. Changes in the relative position of facial markers were calculated using the Euclidean distance between a reference marker, and the relevant marker. The nose bridge acted as the reference marker for lower lip and eyebrow markers, while the sides of the head were used for lip corners. This method is invariant to translational and rotational head movement, isolating motion that is specific to facial expressions. Movement data were then smoothed using a fourth-order zero-phase shift Butterworth filter (cut-off frequency 4 Hz).

Results

Preliminary inspection of the data revealed that movement of two facial features was highly relevant to the experimental conditions: the eyebrows and lip corners. We focused on the analysis of motion capture data from two markers: the middle of the left eyebrow (BROW) and the left lip corner (LLC). Our focus on the left side of the face was motivated by evidence that facial movements are of greater magnitude on the left side because the right hemisphere is dominant for facial expressions (Sackeim, Gur, & Saucy, 1978). As a control, we also analyzed movement of the lower lip (LL). Movement of the lower lip is associated with vocalization but should not vary as a function of the intended emotion of sung phrases.

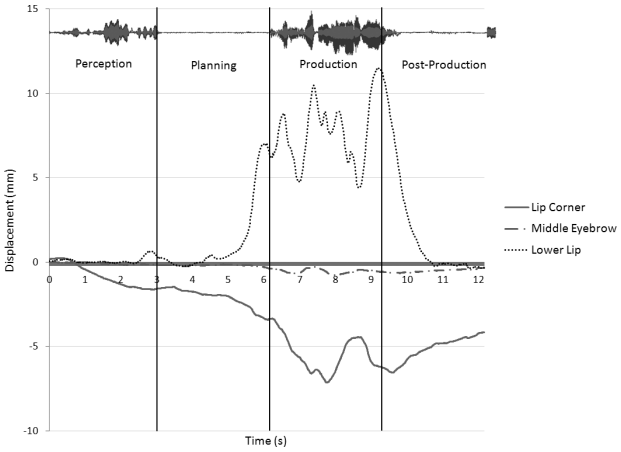

FIGURE 1. Time course of feature displacement in the happy condition, averaged across captures and participants. Time course is divided up into four distinct phases or epochs: perception, planning, production, and post-production. Vocalization is shown in the acoustic signal.

RAW CAPTURES

Figure 1 provides an example of the time course of BROW and LLC averaged across the seven participants (42 captures) of the happy condition. The figure illustrates that during exposure to the target song, the participant exhibited a form of mirroring, or emotional synchronization, with movement of the LLC. The initial displacement of the lip corner was maintained and then increased during the planning phase (epoch 2), suggesting that the emotional synchronization that emerged during the target stimulus also functioned as advanced planning for the imitation phase. Movement of the LLC increased during the production phase (epoch 3) and partially subsided once the participant completed the imitation (epoch 4). Advanced planning also was observed for the lower lip, with movement initiated during the planning epoch and prior to vocalization.

STATISTICAL ANALYSES

The analysis focused on three markers: the middle of the left eyebrow (BROW); the left lip corner (LLC), and

the lower lip (LL). Mean BROW, LLC, and LL displacement values within each epoch were obtained for each emotion and capture, and subjected to ANOVA with repeated-measures on Emotion (3 levels: happy, sad, and neutral), Epoch (4 levels: perception, planning, production, and post-production), and Capture (6 levels: capture trial). A separate ANOVA was conducted for each of the three features. Figure 2 displays mean displacement values for each feature across participants and captures. The sign of each feature's movement in Figure 2 is dependent upon the reference marker used to calculate it. For BROW in Panel A, the displacement is negative because when the eyebrow furrows, the marker moves closer to the nose bridge (displacement decreases). For the LLC in Panel B, the displacement is negative because in the act of smiling, the marker moves closer to the side of the head (displacement decreases). For LL in Panel C, the displacement is positive because when the mouth opens, the marker moves further away from the nose bridge (displacement increases).

For the BROW marker, there were significant main effects of Epoch, $F(3, 18) = 7.06$, $p < .01$, and Emotion, $F(2, 12) = 19.17$, $p < .001$, and a significant interaction between Epoch and Emotion, $F(6, 36) = 12.69$, $p < .001$. There were no significant effects related to Capture.

The main effect of Epoch indicates that eyebrow movement was greater in some epochs than in others. This result is not surprising because greater movement would be expected in the production phase, when participants were actually performing the sung phrase. The main effect of Emotion indicates that eyebrow movement occurred to a greater extent for some emotions over others. In particular, greater eyebrow movement was observed for sad songs than for happy or neutral songs. The significant interaction between Epoch and Emotion illustrates that the time course of facial expressions differed significantly depending on emotional intentions. Whereas there was a gradual increase in the mean displacement of eyebrows across the four epochs for happy and neutral songs, mean displacement values increased dramatically in the imitation phase for sad songs.

To explore the significant interactions between Emotion and Epoch observed for BROW, we conducted an ANOVA for each epoch separately, with repeated measures on Emotion and Capture. The effect of Emotion was not significant in the *perception* epoch, $F(2, 12) = 1.47$, *n.s.*, but it was significant in the *planning*, $F(2, 12) = 4.28$, $p < .05$, *production*, $F(2, 12) = 24.86$, $p < .001$, and *post-production* epochs, $F(2, 12) = 15.22$, $p < .001$. In all four epochs, mean displacement values were greater for the sad condition than for the happy or neutral conditions, driven by a tendency to furrow the eyebrows during sad performances.

For the LLC, there were significant main effect of Epoch, $F(3, 18) = 13.50$, $p < .001$, and Emotion, $F(2, 12) = 59.54$, $p < .001$, and a significant interaction between Epoch and Emotion, $F(6, 36) = 21.37$, $p < .001$. Again, there were no significant effects related to Capture. The main effect of Epoch indicates that movement of the lip corner was greater in some epochs than in others. As with eyebrow movement, this result is not surprising because greater lip movement would be expected during imitation when participants were singing. The main effect of Emotion indicates that movement of the lip corner varied with regard to the emotion expressed. Movement was greater for happy songs relative to neutral songs. The significant interaction between Epoch and Emotion illustrates that the time course of lip movements differs significantly depending on emotional intentions.

To explore the significant interactions between Emotion and Epoch observed for LLC, we conducted an ANOVA for each epoch separately, with repeated measures on Emotion and Capture. The effect of Emotion was significant in the *perception*, $F(2, 12) = 11.03$, $p < .01$, *planning*, $F(2, 12) = 16.62$, $p < .001$, *production*, $F(2, 12) = 82.25$, $p < .001$, and *post-production* epochs, $F(2, 12) = 43.53$, $p < .001$. In all four epochs, mean displacement values were greater for the happy condition than for the sad or neutral condition, with the lip corner raised the highest during happy performances.

For the LL, there were significant mains effect of Epoch, $F(3, 18) = 66.03$, $p < .001$. As anticipated, there was no significant main effect of Emotion, $F(2, 12) = 3.49$, *n.s.*, and no significant interaction between Epoch and Emotion, $F(6, 36) = 2.14$, *n.s.* There were also no significant effects related to Capture. The main effect of Epoch reflects that movement of the lower lip was greater in the imitation epoch (when participants were actually singing) than it was in other epochs.

Experiment 2

Experiment 2 was conducted to supplement the motion capture data in Experiment 1 with another sensitive measure of facial movement: electromyography. The technique was used to examine zygomatic activity and corrugator muscle activity, which are correlated with positive and negative affective states, respectively (Cacioppo, Berntson, Larsen, Poehlmann, & Ito, 2000; Dimberg et al., 2000; Schwartz, Fair, Salt, Mandel, & Klerman, 1976).

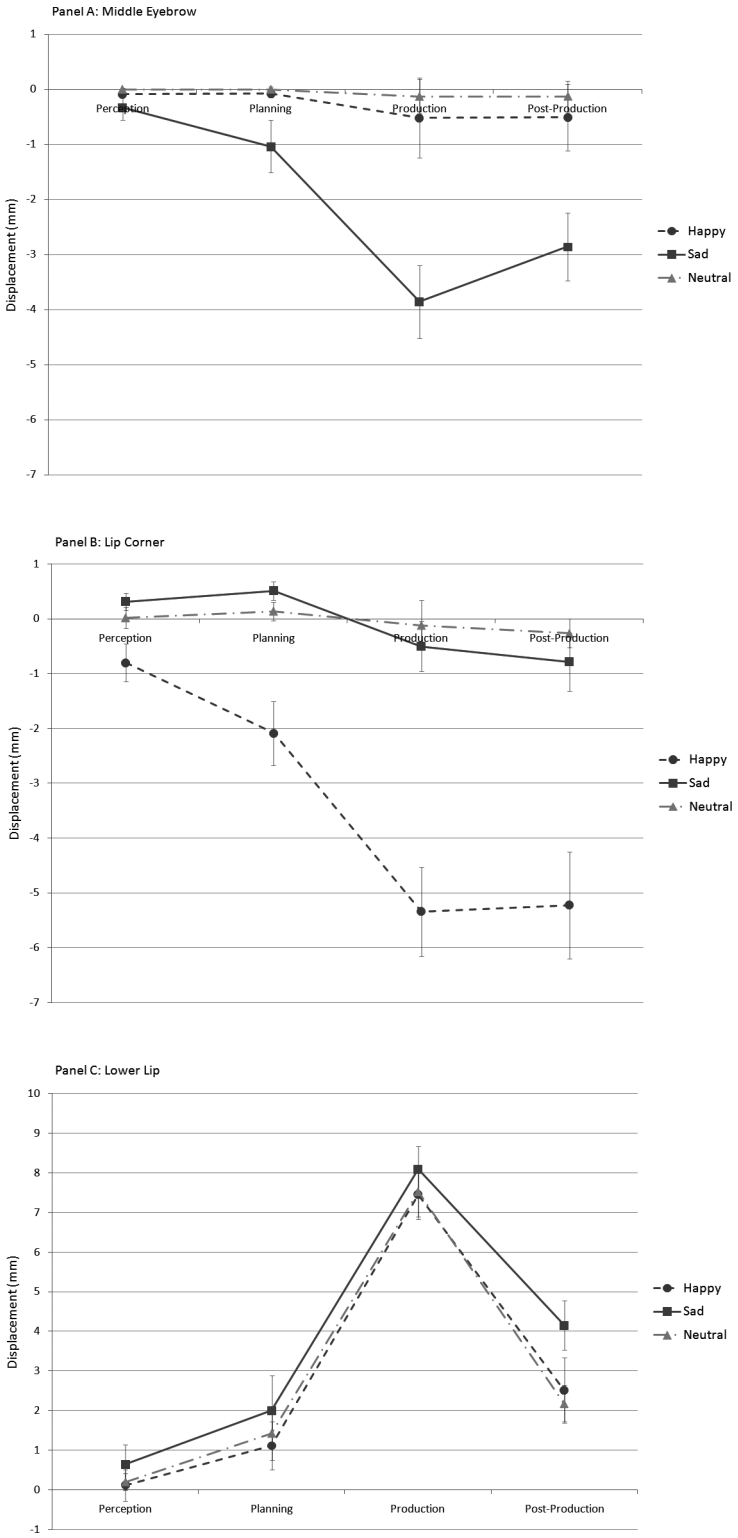

FIGURE 2. Panel A. Mean displacement of the left middle eyebrow in the four epochs. Panel B. Mean displacement of the left lip corner in the four epochs. Panel C. Mean displacement of the lower lip in the four epochs.

Method

PARTICIPANTS

Four musically trained students from Ryerson University acted as participants (3 women and 1 man, $M = 20.75$ years, $SD = 1.26$, range = 19-22). None had participated in Experiment 1. Participants had between 9 and 17 years of music training ($M = 12.5$ years, $SD = 3.70$) and some theatre experience. Testing took approximately 90 minutes, with participants paid $25 CAN for their participation.

STIMULI AND MATERIALS

The presentation stimuli consisted of the same 18 video recorded performances utilized in the motion capture experiment. EMG data were sampled at a rate of 200 Hz using a Biopac EMG100C amplifier and an MP100 data acquisition system under the control of a Mac mini computer. Raw electrical potentials were obtained using 4 mm silver-silver chloride (Ag/AgCl) electrodes and electrode gel applied to the skin using adhesive disks. Electrodes were placed over zygomatic major (smiling) and corrugator supercilli (frowning) muscles on the left side of the face. Stimuli were presented using an Intel-based PC connected to a 19'' TFT LCD monitor and a pair of Logitech X120 Speakers.

Sung phrases were recorded using a Sony Handycam HDR-SR1. The output from the speakers (target stimuli) and the vocal utterances of the participants were captured using a Rode NTK tube condenser microphone. Signal gain was applied using an MBOX1 audio interface, and was output to the Biopac MP100 for synchronisation with EMG data.

PROCEDURE

Participants were seated in a sound attenuated chamber with the viewing screen elevated to eye height. As in the motion capture experiment, participants were asked to observe and reproduce the stimuli, paying particular attention to the emotion being expressed, both visually and aurally. Timing, randomization, and blocking of trials were identical to that described for the motion capture experiment.

MOVEMENT ANALYSIS

Preliminary processing of EMG data involved full-wave rectification (to enable averaging within epochs). To facilitate graphic presentation, data reported in figures also were passed through a fourth-order zero-phase shift Butterworth filter (cut-off frequency = 0.4 Hz).

Results

Figure 3 displays mean EMG activity for zygomatic and corrugator muscles for each emotional intention in the four epochs, averaged across all trials and participants. The correlation between corrugator and zygomatic muscle activity was positive and significant in happy, sad, and neutral production conditions, $r(2432) = .77$, .90, and .71 respectively, $p < .001$. The high correlations are attributable to the common pattern of muscle activation that is necessary to produce the standard melody.

In order to assess muscle activity that can be attributed to emotional expression alone, composite measures were obtained for emotional and neutral productions. For emotional productions, an *emotional composite* measure was obtained by summing together dominant muscle activity across happy and sad conditions; i.e., zygomatic activity for happy production and corrugator activity for sad production. This particular composite measure has previously been used to assess overall facial responsiveness to emotional stimuli (Brown, Bradley & Lang, 2006). For neutral productions, zygomatic and corrugator activity were summed to obtain a *neutral composite*. Figure 4 plots these two composite measures (Butterworth filtered) for the four epochs. Although both the emotional and neutral composite measures are influenced by the act of vocal production, the emotional composite also reflects muscle activity due to emotionality. By comparing these two composite measures, it is possible to dissociate movements related to production from movements related to emotionality. The difference between the two composites may be interpreted as the effect of emotional expression on muscle activity.

STATISTICAL ANALYSIS

For each participant, the mean emotional and neutral composite value within each epoch was obtained for each emotion and capture, and subjected to ANOVA. Because of the low number of participants, *Capture* was treated as the random variable ($n = 6$), and *Emotionality* (2 levels), *Epoch* (4 levels), and *Participant* (4 levels) were treated as repeated measures variables. As illustrated in Figure 5, there was a main effect of Emotionality, $F(1, 5) = 38.54$, $p < .01$, with higher EMG activity for emotional expressions than for neutral expressions, and a main effect of Epoch, $F(3, 15) = 57.68$, $p < .001$, with EMG activity reaching maxima in the production epoch. There was also a significant interaction between Emotionality and Epoch, $F(3, 15) = 10.24$,

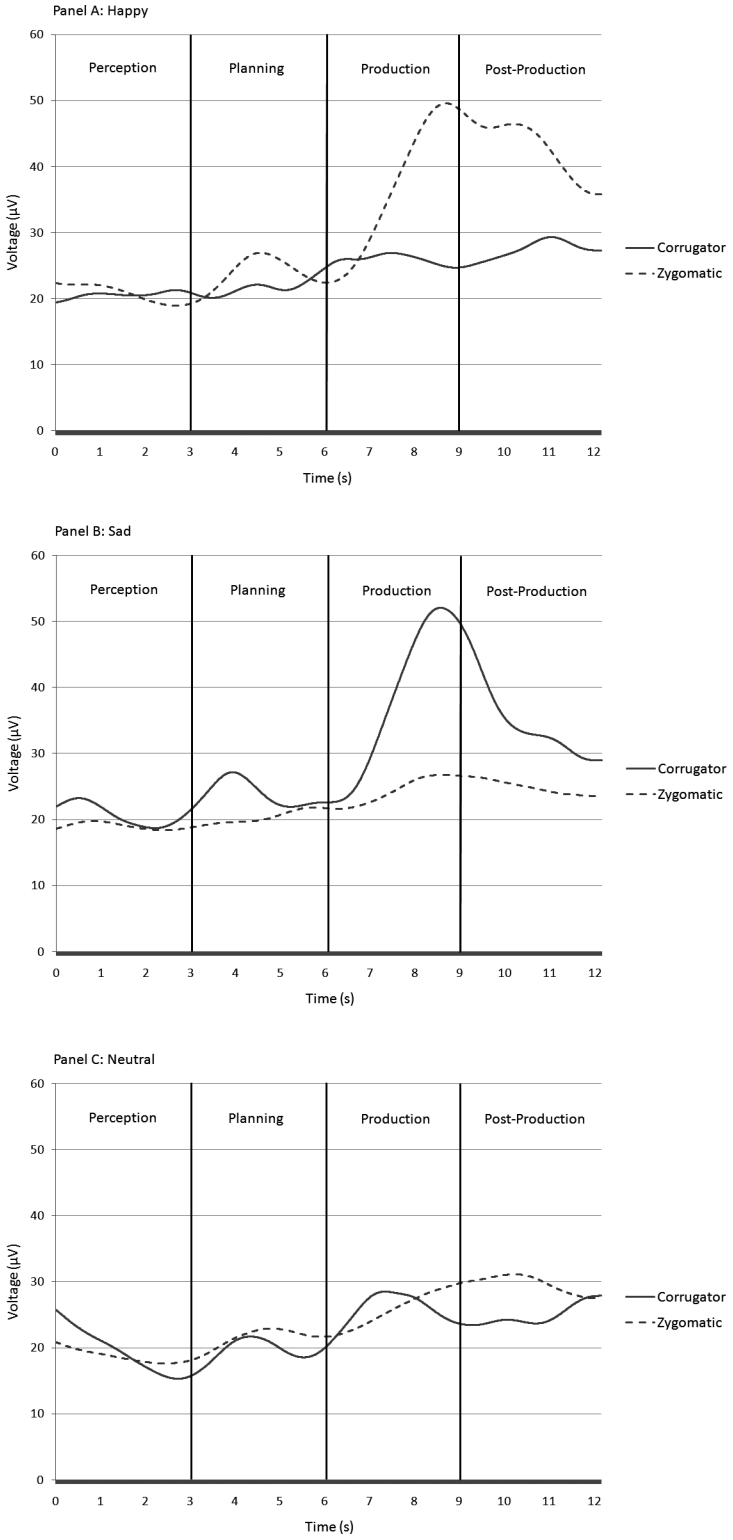

FIGURE 3. EMG activity of zygomatic and corrugator muscles in happy, sad, and neutral conditions.

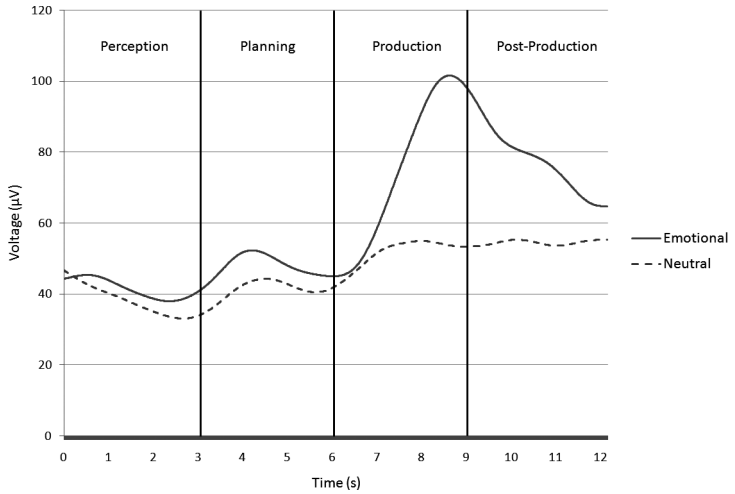

FIGURE 4. Composite EMG functions for emotional and neutral conditions.

$p < .001$. We also observed a significant three-way interaction involving Epoch, Emotion, and Participant, $F(9, 45) = 4.64$, $p < .0001$. This interaction is attributable to variability in the extent of difference between emotional and neutral expressions across epochs and participants. However, consistent with means reported in Figure 5, all participants showed greater EMG activity for emotional expression than for neutral expression.

To further investigate the effects of emotional expression within each epoch, a separate ANOVA was run for each epoch in which Capture was treated as the random variable, with repeated measures on Emotionality (emotional vs. neutral) and Participant. The effect of Emotionality was significant in the *perception*, $F(1, 5) = 8.94$, $p < .05$, *planning* (marginal), $F(1, 5) = 5.24$, $p = .07$, *production*, $F(1, 5) = 20.86$, $p < .01$, and *post-production* epochs, $F(1, 5) = 25.69$, $p < .01$. In all four epochs, the composite EMG was higher for emotional conditions than for neutral conditions.

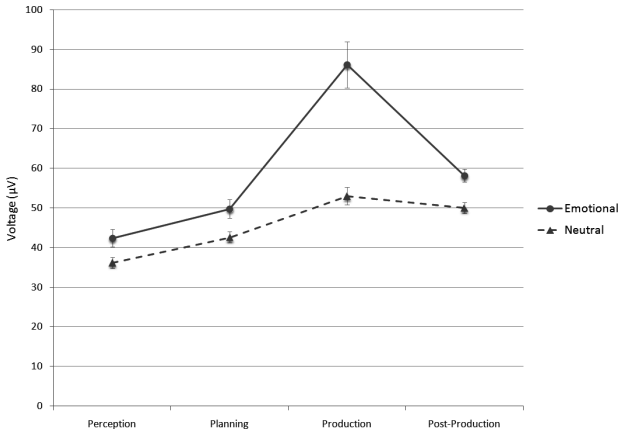

FIGURE 5. Mean emotional and neutral composite EMG values within each epoch.

A second type of composite EMG was next derived from the difference between zygomatic activity in happy productions and corrugator activity in sad productions. A deviation from zero on this composite measure indicates a mismatch in emotion-mediated muscle activity. Positive deviations indicate relatively more activity induced by positive affect (more zygomatic activity) and negative deviations indicate relatively more activity induced by negative affect (more corrugator activity). Figure 6 displays the composite across the four epochs. The largest differences from zero are positive and can be observed in the *post-production* epoch. These differences could be attributable to biophysical differences in the latency of activity across the two muscles. However, a more likely explanation is that singers were more inclined to extend the expression of positive affect beyond the imitation epoch than they were to extend the expression of negative affect.

Discussion

The results of this investigation indicate that facial expressions of emotion supplement the acoustic channel of music in multiple ways. First, facial movements that were tracked by motion capture and EMG were evident in the perception epoch when musicians merely watched another musician sing an emotional melodic phrase. Such movements suggest that perceptual mechanisms

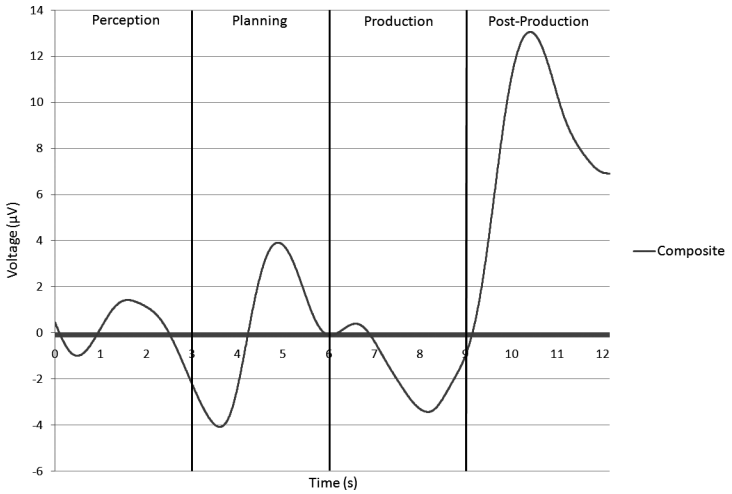

FIGURE 6. Composite difference between zygomatic EMG for happy conditions and corrugator EMG for sad conditions.

engaged processes of motor mimicry and/or emotional synchronization with the sung stimulus, or that participants consciously used facial movements to enhance their experience of the emotion being conveyed.

Second, facial expressions were observed in the planning phase of the presentation stimuli, suggesting that they function for the purpose of motor and emotional planning. Some movement is expected immediately prior to a vocal performance, as a singer must utilize facial muscles in order to produce a vocal utterance. However, movements were observed well in advance of those that would be required for vocal production, suggesting that their function extends beyond production constraints and may play a role in motor and emotional planning.

Third, facial expressions were observed during emotional song production, and differed depending on the intended emotion. These findings corroborate evidence that music performers use facial expressions for emotional communication (Thompson et al., 2005; Thompson et al., 2008). However, they extend this research by identifying the facial features and movements that contribute to this communication. Precise definition of the facial cues to emotion during sung performances will be an important next step in this research, and may inform computational models of emotional communication in music (Livingstone, Muhlberger, Brown, & Loch, 2007; Livingstone, Muhlberger, Brown, & Thompson, 2009).

Fourth, movement and muscle activity were observed in the post-production epoch after participants completed their imitation, indicating that facial expressions of emotion can linger beyond the acoustic dimension of music. Such expressions may function to support and emphasize the emotional message by extending visual cues well beyond the time frame of the acoustic signals of emotion. They also imply that emotional "closure" may occur subsequent to structural closure.

The precise function of movements observed during perception of the model singer cannot be determined from the current data. One possibility is that they do not reflect emotion recognition but represent a general facial response to any type of visual stimulus. Analyses of movement in the perception epoch rules out this interpretation, however: lip corner movement was reliably different in the three emotion conditions, indicating that movements were specific to the emotion perceived. Moreover, informal observation of participants in the perception epoch suggested that they were genuinely attempting to grasp the emotional connotation of the target stimuli by emulating the emotions communicated in the target performance.

Another possibility is that these emotion-specific expressions functioned to enhance emotional identification (Niedenthal et al., 2001; Stel & van Knippenberg, 2008). Enhanced emotional identification, in turn, should lead to more effective emotional imitation during the production phrase. Because participants were explicitly asked to imitate facial expressions, however, it is not possible to conclude that these movements occurred automatically. Participants may have consciously used the perception epoch as an opportunity to practice and prepare for their imitation in the production epoch. Nevertheless, perceivers do spontaneously mimic facial expressions of emotion under similar experimental conditions (e.g., Dimberg, 1990). Moreover, facial mimicry is associated with the mirror

neuron system, which has been implicated in music, theory-of-mind, and empathy (Gallese, 2003; Schulte-Rüther, Markowitsch, Fink, & Piefke, 2007; Sonnby-Borgstrom et al., 2003; Stel, van Baaren, & Vonk, 2008; Watkins, Strafello, & Paus, 2003). By including a perception-only condition, future research may determine whether facial expressions occur spontaneously during the perception of singing.

The picture that is emerging from this research is that facial expressions are related to musical activities in ways that extend well beyond the temporal boundaries of sound production. Emotional facial expressions surround the production of the acoustic signal, providing a foreshadowing of the music and a form of emotional closure that occurs well after the music had ended. Facial expressions also occur during the perception of music, illustrating a form of facial mimicry, or emotional synchronization, which may reflect internal processes such as attention and recognition, and could conceivably implicate the involvement of the mirror-neuron system.

Not all facial movements varied as a function of the emotion conveyed. Considerable movement was observed for the lower lip in the planning, production, and post-production epochs, but this movement was very similar across the three emotions examined, with no effect of emotion reported. That is, emotional facial expressions that occurred outside of vocal production were restricted to specific features such as the lip corner and the eyebrow, while other facial features such as the lower lip were exclusively associated with vocal production and therefore did not reflect emotional communication. Whereas certain facial movements are likely to operate under the control of emotional processes, others operate under the control of processes that are restricted to vocal production.

The emotions examined in this investigation were differentiable by two features: eyebrow and lip corner. More nuanced emotions may require additional features for differentiation, and future research may determine the minimum feature set required to communicate the full range of facial emotions. Cues arising from rigid head motion may be particularly relevant, given their role in communicating structural information in music (Thompson & Russo, 2007) and speech (Yehia, Kuratate, & Vatikiotis-Bateson, 2002).

Evidence of emotion-dependent facial expressions outside of the production epoch was observed in both motion capture and electromyography. In facial EMG research, it is assumed that greater muscle activation leads to greater EMG voltage recorded from the surface of the skin. By comparing facial EMG data with motion capture data, it is possible to connect motion capture results to underlying muscle activation. Our results are consistent with the assumption that corrugator EMG activity was a determinant of eyebrow movement and zygomatic EMG activity was a determinant of lip corner movement.

Taken together, data from motion capture and EMG provide strong support for the notion that facial expressions have multiple functions in music. There is now ample evidence that they are used during music performance and significantly influence the perception of music. The current data suggest that they also function beyond the time frame of sound production. Although the current investigation provides only preliminary data on the matter, it suggests an important avenue for future research.

Author Note

This research was supported by grants awarded to the second author from the Macquarie University Strategic Infrastructure Scheme (MQSIS) and the Australian Research Council (DP0987182), and by a grant from the Natural Sciences and Engineering Research Council of Canada awarded to the third author. We thank Rachel Bennets, Lisa Chan, Lucy Michalewska, George Narroway, Bojan Nekovic, and Lena Quinto for research and technical assistance, and Caroline Palmer for helpful discussions.

Correspondence concerning this article should be addressed to Bill Thompson, Department of Psychology, Macquarie University, Sydney, NSW, 2109, Australia. E-MAIL: Bill.Thompson@mq.edu.au

References

Bigand, E., Filipic, S., & Lalitte, P. (2005). The time course of emotional responses to music. *Annals of the New York Academy of Sciences, 1060*, 429-437.

Blood, A. J., & Zatorre, R. J. (2001). Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion. *Proceedings of the National Academy of Sciences, 98*, 11818-11823.

Brown, L. M., Bradley, M. M., & Lang, P. J. (2006). Affective reactions to pictures of ingroup and outgroup members. *Biological Psychology, 71,* 303-311.

Bush, L. K., Barr, C. L., McHugo, G. J., & Lanzetta, J. T. (1989). The effects of facial control and facial mimicry on subjective reactions to comedy routines. *Motivation and Emotion, 13*, 31-52.

Cacioppo, J. T., Berntson, G. G., Larsen, J. T., Poehlmann, K. M., Ito, T. A. (2000). The psychophysiology of emotion. In M. Lewis & J. M. Haviland-Jones (Eds.), *Handbook of emotions* (pp. 173-191). New York: The Guilford Press.

Chartrand, T. L., & Bargh, J. A. (1999). The chameleon effect: The perception–behavior link and social interaction. *Journal of Personality and Social Psychology, 76*, 893-910.

Dahl, S., & Friberg, A. (2004). Expressiveness of musician's body movements in performances on marimba. In A. Camurri & G. Volpe (Eds.), *Gesture-based communication in human-computer interaction* (Vol. 2915, pp. 479-486). Heidelberg: Springer Berlin.

Dahl, S., & Friberg, A. (2007). Visual perception of expressiveness in musicians' body movements. *Music Perception, 24*, 433-454.

Darwin, C. (1965). *The expression of emotions in man and animals*. Chicago: University of Chicago Press. (Original work published 1872)

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

Davidson, J. W., & Correia, J. S. (2002). Body movement. In R. Parncutt & G. E. McPherson (Eds.), *The science and psychology of music performance* (pp. 237-250). New York: Oxford University Press.

de Gelder, B., Vroomen, J., & Pourtois, G. (1999). Seeing cries and hearing smiles: Crossmodal perception of emotional expressions. In G. Aschersleben, T. Bachmann, & J. Musseler (Eds.), *Cognitive contributions to the perception of spatial and temporal events* (pp. 425-438). Amsterdam: Elsevier Science.

di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., & Rizzolatti, G. (1992). Understanding motor events: A neurophysiological study. *Experimental Brain Research, 91*, 176-80.

Dimberg, U. (1982). Facial reactions to facial expressions. *Psychophysiology, 19*, 643-647.

Dimberg, U., & Lundquist, O. (1988). Facial reactions to facial expressions: Sex differences. *Psychophysiology, 25*, 442-443.

Dimberg, U., Thunberg, M., & Elmehed, K. (2000). Unconscious facial reactions to emotional facial expressions. *Psychological Science, 11*, 86-89.

Ferrari, P. F., Gallese, V., Rizzolatti, G., & Fogassi, L. (2003). Mirror neurons responding to the observation of ingestive and communicative mouth actions in the monkey ventral premotor cortex. *European Journal of Neuroscience, 17*, 1703-1714.

Gallese, V. (2001). The "shared manifold" hypothesis. *Journal of Consciousness Studies, 8*, 33-50.

Gallese, V. (2003). The roots of empathy: The shared manifold hypothesis and the neural basis of intersubjectivity. *Psychopathology, 36*, 171–180.

Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain, 119*, 593-609.

Gallese, V., & Goldman, A. (1998). Mirror neurons and the simulation theory of mindreading. *Trends in Cognitive Sciences, 2*, 493-501.

Geringer, J. M., Cassidy, J. W., & Byo, J. L. (1997). Nonmusic majors' cognitive and affective responses to performance and programmatic music videos. *Journal of Research in Music Education, 45*, 221-233.

Gillespie, R. (1997). Ratings of violin and viola vibrato performance in audio-only and audiovisual presentations. *Journal of Research in Music Education, 45*, 212-220.

Godøy, R. I., Haga, E. & Jensenius, A. R. (2006). Playing 'air instruments': Mimicry of sound-producing gestures by novices and experts. In S. Gibet, N. Courty, & J. F. Kamp (Eds.), *Gesture in human-computer interaction and simulation: 6th international gesture workshop* (pp. 256-267). Berlin: Springer-Verlag.

Goldinger, S. D. (1998). Echos of echoes? An episodic theory of lexical access. *Psychological Review, 105*, 251-279.

Gordon, R. (1986). Folk psychology as simulation. *Mind and Language, 1*, 158–171.

Heal, J. (1986). Replication and functionalism. In J. Butterfield (Ed.), *Language, mind and logic* (pp. 135-150). Cambridge, UK: Cambridge University Press.

Hess, U., & Blairy, S. (2001). Facial mimicry and emotional contagion to dynamic emotional facial expressions and their influence on decoding accuracy. *International Journal of Psychophysiology, 40*, 129-141.

Humphrey, N. K. (1976). The social function of intellect. In P. P. G. Bateson & R. A. Hinde (Eds.), *Growing points in ethology* (pp. 303-321). Cambridge, UK: Cambridge University Press.

Juslin, P. N. (2001). Communicating emotion in music performance: A review and a theoretical framework. In P. N. Juslin & J. A. Sloboda (Eds.), *Music and emotion: Theory and research* (pp. 309-337). Oxford: Oxford University Press.

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

Juslin, P. N., & Västfjäll, D. (2008). Emotional responses to music: The need to consider underlying mechanisms. *Behavioral and Brain Sciences, 31*, 559-621.

Khalfa, S., Peretz, I., Blondin, J.-P., & Manon, R. (2002). Event-related skin conductance responses to musical emotions in humans. *Neuroscience Letters, 328*, 145-149.

Koelsch, S. (2005). Investigating emotion with music: Neuroscientific approaches. *Annals of the New York Academy of Sciences, 1060*, 412-418.

Kohler, E., Keysers, C., Umilta, M. A., Fogassi, L., Gallese, V., & Rizzolatti, G. (2002). Hearing sounds, understanding actions: Action representation in mirror neurons. *Science, 297*, 846-48.

Lahav, A., Saltzman, E., & Schlaug, G. (2007). Action representation of sound: Audiomotor recognition network while listening to newly acquired actions. *Journal of Neuroscience, 27*, 308-314.

Leman, M. (2007) *Embodied music cognition and mediation technology*. Cambridge, MA: MIT Press.

Livingstone, S. R., Muhlberger, R., Brown, A. R., & Loch, A. (2007). Controlling musical emotionality: An affective computational architecture for influencing musical emotions. *Digital Creativity, 18*, 43-53.

Livingstone, S. R., Muhlberger, R., Brown, A. R., & Thompson, W. F. (2009). *Changing musical emotion through score and performance with a computational rule system*. Manuscript submitted for publication.

Livingstone, S. R. & Thompson, W. F. (2009). The emergence of music from the Theory of Mind. *Musicae Scientiae, Special Issue,* 83-115.

McFarland, D. H. (2001). Respiratory markers of conversational interaction. *Journal of Speech, Language, and Hearing Research, 44*, 128-143.

McGurk, H., & MacDonald, J. (1974). Hearing lips and seeing voices. *Nature, 264*, 746-748.

Molnar-Szakacs, I., & Overy, K. (2006). Music and mirror neurons: From motion to 'e'motion. *Social Cognitive and Affective Neuroscience, 1,* 235-241.

Neumann, R., & Strack, F. (2000). "Mood contagion": The automatic transfer of mood between persons. *Journal of Personality and Social Psychology, 79*, 211-223.

Niedenthal, P. M., Brauer, M., Halberstadt, J. B., & Innes-Ker, Å. H. (2001). When did her smile drop? Facial mimicry and the influences of emotional state on the detection of change in emotional expression. *Cognition and Emotion, 15*, 853-864.

Oberman, L. M., & Ramachandran, V. S. (2007). The simulating social mind: The role of the mirror neuron system and simulation in the social and communicative deficits of autism spectrum disorders. *Psychological Bulletin, 133*, 310-327.

Paccalin, C., & Jeannerod, M. (2000). Changes in breathing during observation of effortful actions. *Brain Research, 862*, 194-200.

Palmer, C. (2005). Time course of retrieval and movement preparation in music performance. *Annals of the New York Academy of Sciences, 1060*, 360-367.

Palmer, C. (2006). The nature of memory for music performance skills. In E. Altenmüller, M. Wiesendanger & J. Kesselring (Eds.), *Music, motor control and the brain* (pp. 39-53). Oxford, UK: Oxford University Press.

Pietromonaco, P. R., Zajonc, R. B., & Bargh, J. A. (1981, August). *The role of motor cues in recognition for faces*. Paper presented at the 89th Annual Convention of the American Psychological Association, Los Angeles, CA.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences, 4*, 515-526.

Preston, S. D., & de Waal, F. B. M. (2002). Empathy: Its ultimate and proximate bases. *Behavioral and Brain Sciences, 25*, 1-72.

Radeau, M., & Bertelson, P. (1974). The after-effects of ventriloquism. *Quarterly Journal of Experimental Psychology, 26*, 63-71.

Rickard, N. S. (2004). Intense emotional responses to music: A test of the physiological arousal hypothesis. *Psychology of Music, 32*, 371-388.

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience, 27*, 169-192.

Rosenblum, L. D. (1988). An audiovisual investigation of the loudness/effort effect for speech and nonspeech events. *Journal of the Acoustical Society of America, 84*, S159.

Sackeim, H. A., Gur, R. C., & Saucy, M. C. (1978). Emotions are expressed more intensely on the left side of the face. *Science, 202*, 434-436.

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

Schubert, E. (2001). Continuous measurement of self-report emotional response to music. In P. N. Juslin & J. A. Sloboda (Eds.), *Music and emotion: Theory and research* (pp. 393-414). Oxford: Oxford University Press.

Schulte-Rüther, M., Markowitsch, H. J., Fink, G. R., & Piefke, M. (2007). Mirror neuron and theory of mind mechanisms involved in face-to-face interactions: A functional magnetic resonance imaging approach to empathy. *Journal of Cognitive Neuroscience, 19*, 1354-1372.

Schutz, M. (2008). Seeing music? What musicians need to know about vision. *Empirical Musicology Review, 3*, 83-108.

Schutz, M., & Lipscomb, S. (2007). Hearing gestures, seeing music: Vision influences perceived tone duration. *Perception, 36*, 888-897.

Schwartz, G. E., Fair, P. L., Salt, P., Mandel, M. R., & Klerman, G. R. (1976). Facial muscle patterning to affective imagery in depressed and nondepressed subjects. *Science, 192*, 489-491.

Sonnby-Borgstrom, M., Jonsoon, P., & Svensson, O. (2003). Emotional empathy as related to mimicry reactions at different levels of information processing. *Journal of Nonverbal Behavior, 27*, 3-23.

Stel, M., van Baaren, R. B., & Vonk, R. (2008). Effects of mimicking: Acting prosocially by being emotionally moved. *European Journal of Social Psychology, 38*, 965-976.

Stel, M., & van Knippenberg, A. (2008). The role of facial mimicry in the recognition of affect. *Psychological Science, 19*, 984-985.

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., Livingstone, S. R. & Palmer, C. (2008, May). *The emotional-production synchronization paradigm: Analysis of time-varying facial signals of emotion*. Paper presented at the Musical Movement and Synchronization Workshop, Leipzig, Germany.

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

Thompson, W. F., Russo, F. A., & Quinto, L. (2008). Audio-visual integration of emotional cues in song. *Cognition and Emotion, 22*, 1457-1470.

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.

Wallbott, H. G. (1991). Recognition of emotion from facial expression via imitation? Some indirect evidence for an old theory. *British Journal of Social Psychology, 30*, 207-219.

Wanderley, M. M. (2002). Quantitative analysis of non-obvious performer gestures. In I. Wachsmuth & T. Sowa (Eds.), *Gesture and sign language in human-computer interaction* (pp. 241-253). Berlin: Springer-Verlag.

Wanderley, M. M., Vines, B. W., Middleton, N., McKay, C., & Hatch, W. (2005). The musical significance of clarinetists' ancillary gestures: An exploration of the field. *Journal of New Music Research, 34*, 97-113.

Watkins, K. E., Strafella, A. P., & Paus, T. (2003) Seeing and hearing speech excites the motor system involved in speech production. *Neuropsychologia, 41*, 989-94.

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