Facial expressions of singers influence perceived pitch relations

William Forde Thompson
Macquarie University, Sydney, New South Wales, Australia

Frank A. Russo
Ryerson University, Toronto, Ontario, Canada

AND

Steven R. Livingstone
McGill University, Montreal, Quebec, Canada

In four experiments, we examined whether facial expressions used while singing carry musical information that can be "read" by viewers. In Experiment 1, participants saw silent video recordings of sung melodic intervals and judged the size of the interval they imagined the performers to be singing. Participants discriminated interval sizes on the basis of facial expression and discriminated large from small intervals when only head movements were visible. Experiments 2 and 3 confirmed that facial expressions influenced judgments even when the auditory signal was available. When matched with the facial expressions used to perform a large interval, audio recordings of sung intervals were judged as being larger than when matched with the facial expressions used to perform a small interval. The effect was not diminished when a secondary task was introduced, suggesting that audio–visual integration is not dependent on attention. Experiment 4 confirmed that the secondary task reduced participants' ability to make judgments that require conscious attention. The results provide the first evidence that facial expressions influence perceived pitch relations.

There is behavioral, cognitive, and neurological evidence that visual information can reinforce or modify auditory experience, leading to the ventriloquism effect (Radeau & Bertelson, 1974) and the McGurk effect (McGurk & MacDonald, 1976). When visual and auditory recordings of speech are manipulated to conflict with one another, the perceptual result is often a compromise. When visual and auditory speech information is reinforcing (as in normal speech), availability of the visual channel improves intelligibility (Middleweerd & Plomp, 1987; Sumby & Pollack, 1954).

Until recently, researchers had rarely considered the effects of visual information on music perception. These effects need not be equivalent to those observed for speech. Musical and linguistic abilities are characterized as distinct cognitive modules (Peretz & Coltheart, 2003) and may recruit different forms of auditory processing in the left and right hemispheres (Zatorre, Belin, & Penhune, 2002). Whether the two domains are associated with similar processes of audio–visual integration has yet to be determined.

Thompson, Graham, and Russo (2005) observed that singers' facial expressions often convey emotion. Emotional facial movements are observed prior to, during, and after the vocal production of a sung phrase (Living-

stone, Thompson, & Russo, 2009). Facial expressions of singers also reflect musical structure. Thompson and Russo (2007) found that facial expressions reflect the size of sung melodic intervals. Participants observed silent videos of musicians singing 13 melodic intervals and judged the size of each interval that the singer was imagined to be singing. Participants could discriminate intervals on the basis of visual information alone. Facial and head movements were correlated with the size of sung intervals.

The present investigation was conducted to explore the latter findings. First, although movement analysis revealed correlations between facial or head movements and interval size, it was unclear which movements influenced judgments. The significance of head movements has been demonstrated for speech perception (Munhall, Jones, Callan, Kuratate, & Vatikiotis-Bateson, 2004), but no study has demonstrated that head movements influence perceived pitch relations. In Experiment 1, we examined interval discrimination under full-view conditions and with facial features occluded. If discrimination of intervals occurs with facial features occluded, it would suggest that head movements provide reliable information about interval size. If discrimination is reduced or eliminated with facial features occluded, it would suggest

W. F. Thompson, bill.thompson@mq.edu.au

that facial features provide additional information about interval size.

A second question, which we addressed with Experiments 2 and 3, concerns whether facial expressions influence the perception of melodic intervals when auditory cues are available. Audio–visual recordings of performances were edited such that the same melodic intervals presented aurally were synchronized with facial expressions used when large and small intervals were sung. Synchronized performances were then presented to participants, who judged the size of the interval.

A third question concerns whether auditory and visual signals are consciously combined, or whether integration occurs preattentively. Participants in Experiments 2 and 3 judged interval size while completing a demanding secondary task. If integration of auditory and visual signals required conscious attention, the presence of a secondary task should reduce integration and, hence, the influence of facial expressions. Finally, Experiment 4 confirmed that the secondary task genuinely occupied attentional resources, interfering with those tasks that do require attention.

EXPERIMENT 1

Do facial and head movements of singers carry information about pitch relations? Three vocalists were recorded singing four ascending melodic intervals. Motion capture was used to examine their facial and head movements. Participants saw the silent video recordings and judged the size of the interval that they imagined the performer to be singing. Judgments were made under conditions in which the face and head were visible (no occlusion) or in which the face was occluded such that only head movements were visible. If facial and head movements collectively carry information about the size of melodic intervals, judgments of interval size under the no-occlusion condition should differ across intervals. If head movements alone carry information about the size of melodic intervals, judgments of pitch distance under the occlusion condition should also differ across the four intervals.

Method

Participants. We recruited 20 participants (19 females, 1 male; mean age = 21.60 years, SD = 1.76, range = 18–49; mean years of music training = 5.0, SD = 1.31, range = 0–16). No participant reported abnormal hearing.

Stimuli and Materials. Three trained vocalists sang ascending melodic intervals of 0, 6, 7, and 12 semitones (unison, augmented fourth, perfect fifth, and octave, respectively). Each interval was sung twice, beginning on each of three pitches: C4, B♭3, and D4. This procedure resulted in 12 sung intervals per singer (4 intervals, 3 starting pitches). Singers practiced each interval before being recorded. During recording, accuracy was reinforced with piano tones presented over Sennheiser HD 555 headphones with tone durations set to 1.5 sec. Singers were asked to sing in a "natural" manner, without compromising accuracy. Performances were recorded using a Sony Handycam HDR-SR1 and an external Sony ECMHST1 electoret condenser microphone. Recordings were edited using Final Cut software.

Performances were highly accurate (within 20 cents of the interval size for all intervals; 1 cent = 0.01 semitone). The 5-sec videos were displayed on a 21-in. Apple CRT display (1,280 × 1,024 pixels) under no-occlusion (full view) and occlusion (face occluded) conditions. For the no-occlusion condition, participants had full view of the singers from the shoulders up. For the occlusion condition, an opaque gray shape was superimposed over the singer's face. The shape moved dynamically with the face, leaving the outline of the head and hair visible. The occlusion conditions were randomized. There were 72 trials (three singers, four intervals, three starting pitches, two occlusion conditions). An additional 72 trials involving different occlusion conditions were randomly interspersed among the trials described above; however, discussion of these trials has been excluded for the sake of brevity.

Facial movements were recorded in a separate session with a Vicon motion-capture camera (4× MX-F20 2-megapixel, MX Ultranet HD, frame rate = 200 Hz). Thirteen markers were placed on each singer's face: three 9-mm-diameter spherical markers (forehead and left and right sides of the head) and ten 4-mm-diameter hemispherical markers (inner and middle of each eyebrow, nose bridge, nose tip, upper and lower lip, and left and right lip corners). Motion capture occurred 15 min after stimulus recording, using a procedure identical to that used for stimulus creation.

Procedure. Participants watched each video and rated the size of the interval they imagined the performer to be singing on a scale from 1 to 7, with 1 = *very small* and 7 = *very large*.

Results

An ANOVA with repeated measures on interval (four intervals) and occlusion (full view, occluded face) revealed a main effect of interval [$F(3,57)$ = 114.89, p < .0001, η_p^2 = .86]. Figure 1 shows means and standard errors for each interval and occlusion condition. For the no-occlusion condition, each increase in interval size (0–6, 6–7, 7–12 semitones) led to a reliable increase in mean ratings of interval size [$t(19)$ = 11.35, 4.14, 2.88; ps < .01; d = 2.19, 0.57, 0.54]. For the occlusion condition, only the 6- and 7-semitone intervals were not discriminated [$t(19)$ = 1.57, n.s.]. Ratings were higher for the 6- than for the 0-semitone interval [$t(19)$ = 9.029, p < .01, d = 0.20] and higher for the 12- than for the 7-semitone interval [$t(19)$ = 2.77, p < .05, d = 0.71]. Thus, visual information arising from the head and face provided reliable signals of interval size, with increased discrimination when facial features were visible.

A significant interaction between interval and occlusion confirmed that discrimination was affected by facial occlusion [$F(3,57)$ = 10.59, p < .0001, η_p^2 = .36]. For the 0-semitone interval, ratings were higher for the occlusion than for the no-occlusion condition [$F(1,19)$ = 18.60, p < .001, η_p^2 = .50]. For the 7- and 12-semitone intervals, ratings were lower for the occlusion than for the no-occlusion condition [$F(1,19)$ = 6.48 and 3.88, p = .02 and .06, η_p^2 = .25 and .17]. This pattern of results indicates greater discrimination of intervals when facial features were available than when only head movements were available.

To corroborate this result, we converted each participant's set of interval size ratings into a single *discrimination score*, calculated as the SD of the mean ratings for the four intervals. A discrimination score of 0 indicates that mean ratings were identical for the four intervals. Discrimination scores were subjected to an ANOVA with repeated measures on singer and occlusion. The effect of singer was not significant [$F(2,38)$ = 2.67, n.s.], nor was
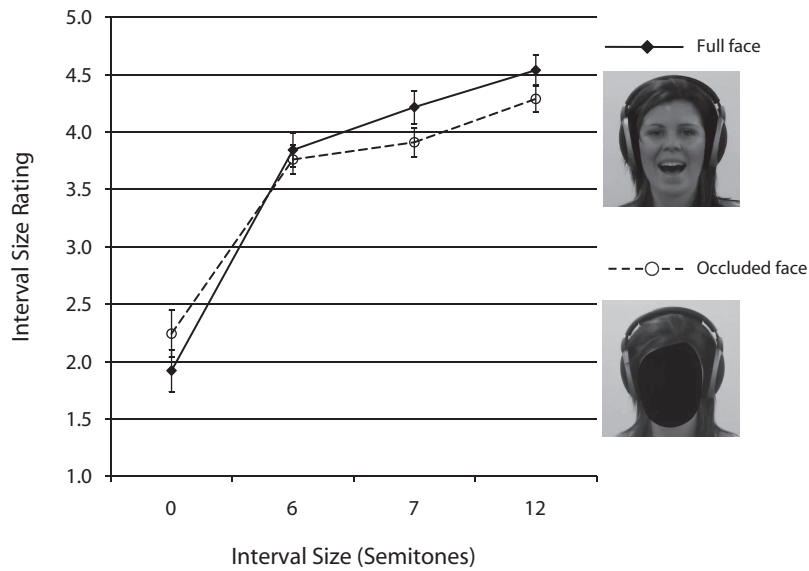
Figure 1. Mean rating of interval size for full face and occluded face conditions. Participants rated pitch intervals on a scale of 1–7. Vertical bars represent standard errors.

the interaction between singer and occlusion [$F(2,38) = 2.35$, n.s.]. However, a significant effect of occlusion revealed that interval discrimination was poorer for occluded ($M = 1.57$, $SD = 0.39$) than for visible ($M = 1.86$, $SD = 0.48$) facial features [$F(1,19) = 15.19$, $p < .001$, $\eta_p^2 = .44$].

Motion-capture data. Raw capture data were reconstructed using Vicon Nexus 1.3.109, with missing data interpolated with spline-curve fitting. Motion data were smoothed by functional data analysis and synchronized with audio output. We computed the maximal displacement of the head, mouth opening, and eyebrow for each interval. Initial eyebrow and mouth opening positions were calculated as the Euclidean distance from the inner left eyebrow to the forehead and from the upper to the lower lip, respectively. The initial head position was calculated as the height of the nose tip marker above the floor. Maximal displacement was calculated as the peak displacement during the second note relative to the marker position prior to production of the first note (singer at rest). Figure 2 illustrates that maximum displacement of the eyebrows, mouth opening, and head increased with interval size. Thus, the movements of singers carry multiple and redundant signals about melodic structure.
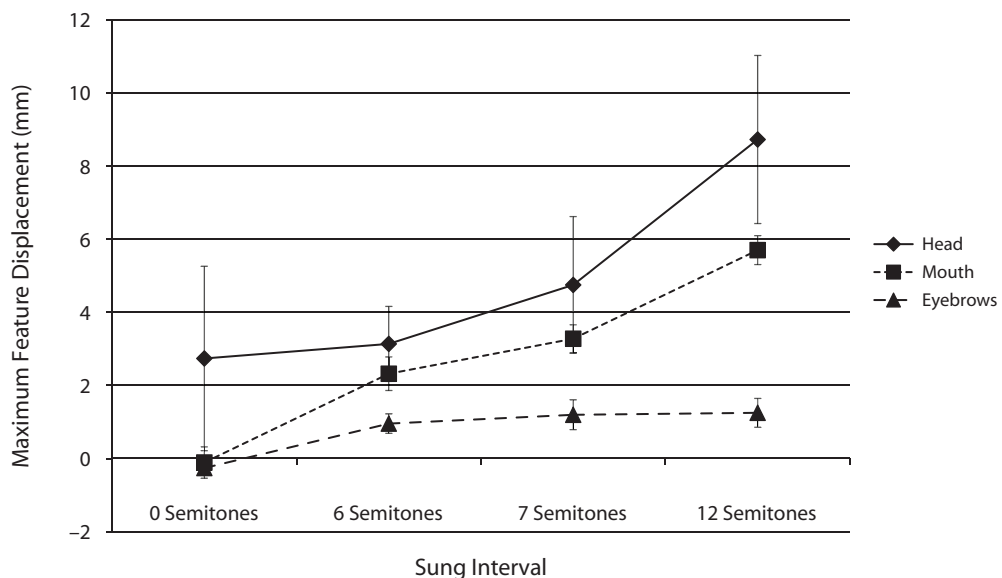
Figure 2. Mean maximum displacement (in millimeters) of head, mouth opening, and eyebrows across singers and starting pitch conditions. Vertical bars represent standard errors.

EXPERIMENT 2

Do facial expressions have an impact when auditory information is available? In Experiment 2, audio and video tracks from separate recordings were synchronized in a congruent (reinforcing) or incongruent (conflicting) manner and presented to listeners. If listeners integrate visual information with the auditory signal, interval size judgments should reflect a compromise between these channels.

While assessing interval size, participants performed a secondary task that involved counting translucent 0s from a succession of 1s and 0s that appeared over the performer's face. Two levels of difficulty were implemented, determined by the speed at which the digits appeared. If integration of audio–visual information requires conscious attention, placing demands on attentional resources by introducing a secondary task should diminish the influence of facial expressions on judgments (Thompson, Russo, & Quinto, 2008; Vroomen, Driver, & de Gelder, 2001). If audio–visual integration occurs automatically, introducing a secondary task should have no effect.

Method

Participants. We recruited 30 participants (28 female, 2 male; mean age = 23.50 years, SD = 7.80, range = 18–49; mean years of music training = 4.57, SD = 5.82, range = 0–16). No participant reported abnormal hearing.

Stimuli and Materials. Presentations were created from audio and video recordings of a musician singing each of four ascending intervals: 0, 6, 7, and 12 semitones. Using Final Cut software, we synchronized sung intervals of two sizes (6 and 7 semitones) with facial expressions used to sing a large (12-semitone) and small (0-semitone) interval. This procedure resulted in four clips.

For each condition of task demand (single- or dual-task conditions), a sequence of 0s and 1s was superimposed over the singer's face during the performance. One, two, or three 0s were flashed in random serial positions. Digits were presented at two rates to manipulate the difficulty of the secondary task: slow (700 msec per digit) or fast (300 msec per digit). Conditions were blocked by task demand and digit speed. The order of conditions was counterbalanced across participants. Half of the participants received dual-task conditions as Blocks 1 and 2; the rest received dual-task conditions as Blocks 3 and 4.

Audio and video recordings were digitized, edited, and presented under the control of a Macintosh Pro (OS X 10.4.11). Videos were displayed on a 21-in. Apple CRT display (1,280 × 1,024 pixels). Audio was presented through Sennheiser HD 555 headphones.

Procedure. Participants rated interval size on a scale from 1 to 7. They were told that digits would appear on the singer's face. In the dual-task condition, they first reported the number of 0s that appeared during the clip (secondary task) and then rated the size of the sung interval (primary task). In the single-task condition, they ignored the digits and focused on rating interval size.

Results

Ratings for the primary task were subjected to an ANOVA with repeated measures on audio interval (6 or 7 semitones), visual interval (0 or 12 semitones), task demand (single or dual task), and digit speed (slow or fast). Ratings were higher when the audio interval was 7 semitones (M = 4.40, SD = 0.95) than when it was 6 semitones (M = 3.13, SD = 0.96) [$F(1,29)$ = 56.78, $p < .001$, η_p^2 = .66], confirming that participants discriminated interval size on the basis of auditory input. Nonetheless, ratings
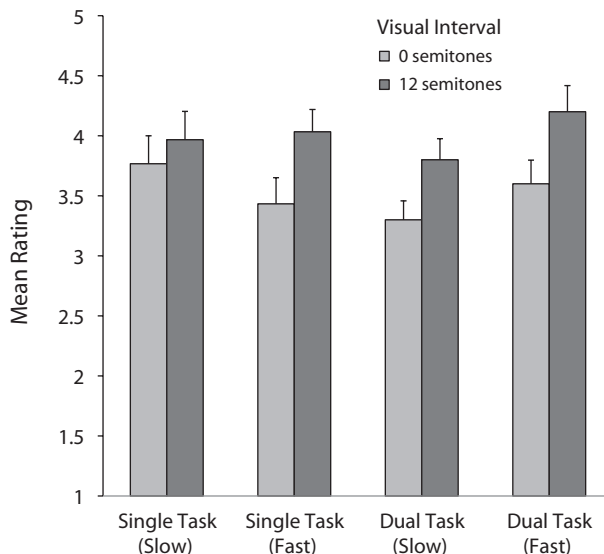
Figure 3. Mean ratings collapsed across the two sung (audio) intervals that were combined with facial expressions used to produce small (0-semitone) and large (12-semitone) intervals. Vertical bars represent standard errors.

were higher when sung intervals were paired with facial expressions used to perform a large (M = 4.00, SD = 0.93) than a small (M = 3.53, SD = 0.85) interval [$F(1,29)$ = 17.53, $p < .001$, η_p^2 = .38]. As is shown in Figure 3, even when the auditory signal was available, facial expressions influenced perceived pitch relations.

A nonsignificant interaction between visual interval and task demand suggested that the influence was independent of attention [$F(1,29) < 1$, n.s.]. There were no effects related to task demand or digit speed. The effect of visual interval was observed even at the most difficult level of the secondary task [$F(1,29)$ = 9.03, $p < .01$, η_p^2 = .24], suggesting that audio–visual integration of sung materials occurs preattentively.

Examination of secondary task performance revealed high accuracy for slow (M = .78, SD = .18) and fast (M = .80, SD = .18) digit rates. Accuracy was similar in the two conditions, implying that participants maintained accuracy levels by allocating greater attentional resources to the fast than to the slow condition.

EXPERIMENT 3

Experiment 2 confirmed that visual information can influence the perception of interval size even when auditory cues are available, and that audiovisual integration occurs preattentively. Two limitations of Experiment 2 motivated a third experiment. First, data were based on a single singer, and corroboration with an additional singer would strengthen conclusions. Second, the sounded intervals used in Experiment 2 differed by only 1 semitone (augmented fourth and perfect fifth; i.e., 6 and 7 semitones), whereas visual intervals were highly contrasting (0 and 12 semitones). Visual influences might not occur if differences in visual intervals are decreased, and differ-

ences in auditory intervals are increased. Experiment 3 was designed to evaluate this possibility and corroborate the results of Experiment 2 using another singer.

Method

Participants. We recruited 18 students (6 female, 12 male; mean age = 19.56, $SD = 0.78$, range = 18–32 years; mean years of music training = 1.18, $SD = 0.33$, range = 0–4). No participant reported abnormal hearing or was involved in Experiment 1 or 2.

Stimuli and Materials. Stimuli were presented on a Macintosh LCD video display with Sennheiser HD-280 headphones. Presentations were created from audio and video recordings of a musician—different from the singer used for Experiment 2—singing three ascending intervals: 2, 7, and 9 semitones. Using Final Cut software, we synchronized sung intervals of two sizes (7 and 9 semitones) with facial expressions used to sing a large (7- or 9-semitone) or small (2-semitone) interval. Sung intervals were never paired with facial expressions used to produce the same interval. Four additional exemplars of each condition were created using ProTools software by pitch shifting the original sung interval up or down by 1 or 2 semitones, yielding five starting pitch positions. There were 20 clips in total (two audio intervals × two visual intervals × five starting positions).

During each performance, a sequence of flashing 0s and 1s was superimposed over the singer's face, as described in Experiment 2. Conditions were blocked by task demand (single or dual task) and digit speed (300 or 700 msec per digit). The order of the four conditions was counterbalanced across participants. Half of the participants received dual task conditions in Blocks 1 and 2; the other half received dual task conditions in Blocks 3 and 4.

Procedure. The procedure was identical to that in Experiment 2.

Results

Ratings were subjected to ANOVAs with repeated measures on audio interval (7 or 9 semitones), visual interval (large or small, defined above), task demand (single or dual task), and digit speed (slow or fast). Ratings were higher when sung intervals were paired with facial expressions used to perform a large ($M = 3.30$, $SD = 0.37$) than a small ($M = 3.08$, $SD = 0.47$) interval [$F(1,16) = 5.49$, $p < .05$, $\eta_p^2 = .26$]. All interactions with visual interval were nonsignificant [$Fs(3,51) < 1$, n.s.], confirming that the effect of visual interval did not depend on attention. The effect of visual interval was observed at the most difficult level of the secondary task [$F(1,17) = 5.16$, $p < .05$, $\eta_p^2 = .23$], suggesting that audio–visual integration of sung materials occurs preattentively.

EXPERIMENT 4

Experiments 3 and 4 indicated that the influence of facial expressions on perceived interval size is unaffected by a secondary task, implying the presence of both automatic and unconscious audio–visual integration. However, this conclusion rests on the assumption that the secondary task genuinely had the capacity to pull attentional resources away from another (primary) task. Experiment 4 tested this assumption. Participants with a range of music backgrounds classified intervals while performing the secondary task. Untrained listeners were trained to classify intervals prior to the experiment. Explicit classification of intervals requires the retrieval of verbal labels, and attention is required to map perceptual input onto mental representations of interval categories, such as perfect fourth (5 semitones), per-

fect fifth (7 semitones), octave (12 semitones), and unison (0 semitones). If the secondary task demands attention, it should interfere with the classification task.

Method

Participants. We recruited 10 participants (7 female, 3 male; mean age = 23.5, $SD = 2.01$, range = 21–27 years; mean years of music training = 4.76, $SD = 1.51$, range = 0–12).

Stimuli and Materials. Stimuli were drawn from recordings used in Experiments 1 and 3, including audio–visual recordings of sung intervals of 6, 7, and 9 semitones produced by three singers. We created multiple exemplars by pitch shifting the original interval up and down by 1 and 2 semitones (ProTools Software). During each performance, digits were flashed over the singer's face, as described in Experiment 2. Conditions were blocked by task demand (single or dual task) and digit speed (300 or 700 msec per digit). The order of conditions was counterbalanced across participants. Half of the participants received dual-task conditions in Blocks 1 and 2; the other half received them in Blocks 3 and 4.

Procedure. Participants classified intervals using a forced choice response: augmented fourth (6 semitones), perfect fifth (7 semitones), and major sixth (9 semitones). Before commencing the experiment, participants received practice trials involving audio-alone presentation of test intervals. Feedback was provided until participants achieved a minimum of 66% accuracy.

All intervals were presented congruently (no manipulation of the original recording). For the single-task condition, participants ignored the digits and focused attention on classifying each interval. For the dual-task conditions, participants reported the number of 0s that appeared and then classified the interval.

Results

An ANOVA with repeated measures on audio interval (6, 7, or 9 semitones), task demand (single or dual task), and digit speed (slow or fast) revealed a main effect of task demand [$F(2,18) = 30.19$, $p < .0001$, $\eta_p^2 = .77$]. Planned contrasts revealed that performance was better in the single-task ($M = 69.90$, $SD = 6.26$) than in the dual-task slow ($M = 64.30$, $SD = 9.75$) condition [$F(1,9) = 21.32$, $p < .0001$, $\eta_p^2 = .70$], which, in turn, was better than in the dual-task fast ($M = 61.60$, $SD = 9.98$) condition [$F(1,9) = 19.24$, $p < .001$, $\eta_p^2 = .68$]. Thus, the secondary task interfered with the primary task (interval classification), and the degree of interference was affected by the rate of presentation. These results confirm that the secondary counting task employed in Experiments 2 and 3 occupied attention.

DISCUSSION

Facial expressions carry information about pitch relations that can be read by viewers and that influence the perception of music. Even when auditory information was available, visual information still influenced judgments. This finding is intriguing, because melodic intervals are defined as auditory events, so visual information should be irrelevant. The effects were undiminished when attention was occupied by a secondary task, suggesting that audio–visual integration occurs automatically and preattentively (Thompson et al., 2008). In that pitch relations are fundamental to musical structure and are evaluated early in processing, the findings illustrate that facial expressions are highly relevant to the perception of music.

During normal face-to-face conversations, eyebrow and head movements reinforce prosodic information (tone of voice), including information about which word in a sentence received emphatic stress and whether a sentence is a statement or question (Bernstein, Demorest, & Tucker, 2000). Our findings indicate that facial movements are similarly important in communicating information about musical pitch.

Facial and head movements may reflect pitch relations for several reasons. First, performers might directly communicate pitch relations through conscious or unconscious movements of facial features, such as the eyebrows, mouth opening, and head. By mapping the extent of pitch change onto observable movements, performers might reinforce the size of the interval and facilitate melodic processing. Such movements may also convey to listeners that pitch changes are intentional. Second, facial expressions may communicate an emotional interpretation of the interval. Larger intervals are generally associated with higher degrees of emotional intensity, which may be reflected in greater movement. Third, performers may inadvertently move their eyebrows and head in response to an arousal state associated with pitch movement. Scherer (2003) observed that increased vocal pitch range is associated with heightened emotional arousal. Similarly, people are more expressive in their visual prosody during heightened emotional states. Thus, performing a large pitch interval may suggest heightened arousal that, in turn, is reflected in face and head movements. Finally, facial and head movements may be introduced to optimize vocal production. Accurate performance of melodic intervals requires rapidly repositioning the vocal apparatus, with larger changes in pitch requiring greater degrees of repositioning.

AUTHOR NOTE

Thompson, Department of Psychology, Macquarie University, Sydney, NSW 2109, Australia (e-mail: bill.thompson@mq.edu.au).

REFERENCES

BERNSTEIN, L. E., DEMOREST, M. E., & TUCKER, P. E. (2000). Speech perception without hearing. *Perception & Psychophysics*, **62**, 233-252.

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*, **26**, 475-488. doi:10.1525/mp.2009.26.5.475

McGURK, H., & MACDONALD, J. (1976). Hearing lips and seeing voices. *Nature*, **264**, 746-748. doi:10.1038/264746a0

MIDDLEWEERD, M. J., & PLOMP, R. (1987). The effect of speechreading on the speech-reception threshold of sentences in noise. *Journal of the Acoustical Society of America*, **82**, 2145-2147. doi:10.1121/1.395659

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. *Psychological Science*, **15**, 133-137. doi:10.1111/j.0963-7214.2004.01502010.x

PERETZ, I., & COLTHEART, M. (2003). Modularity of music processing. *Nature Neuroscience*, **6**, 688-691. doi:10.1038/nn1083

RADEAU, M., & BERTELSON, P. (1974). The after-effects of ventriloquism. *Quarterly Journal of Experimental Psychology*, **26**, 63-71. doi:10.1080/14640747408400388

SCHERER, K. R. (2003). Vocal communication of emotion: A review of research paradigms. *Speech Communication*, **40**, 227-256. doi:10.1016/S0167-6393(02)00084-5

SUMBY, W. H., & POLLACK, I. (1954). Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America*, **26**, 212-215. doi:10.1121/1.1907309

THOMPSON, W. F., GRAHAM, P., & RUSSO, F. A. (2005). Seeing music performance: Visual influences on perception and experience. *Semiotica*, **156**, 203-227. doi:10.1515/semi.2005.2005.156.203

THOMPSON, W. F., & RUSSO, F. A. (2007). Facing the music. *Psychological Science*, **18**, 756-757. doi:10.1111/j.1467-9280.2007.01973.x

THOMPSON, W. F., RUSSO, F. A., & QUINTO, L. (2008). Audio–visual integration of emotional cues in song. *Cognition & Emotion*, **22**, 1457-1470. doi:10.1080/02699930701813974

VROOMEN, J., DRIVER, J., & DE GELDER, B. (2001). Is cross-modal integration of emotional expressions independent of attentional resources? *Cognitive, Affective, & Behavioral Neuroscience*, **1**, 382-387. doi:10.3758/CABN.1.4.382

ZATORRE, R. J., BELIN, P., & PENHUNE, V. B. (2002). Structure and function of auditory cortex: Music and speech. *Trends in Cognitive Sciences*, **6**, 37-46. doi:10.1016/S1364-6613(00)01816-7