Emotion
2016, Vol. 16, No. 3, 365–380

1528-3542/16/$12.00 http://dx.doi.org/10.1037/emo0000106

Head Movements Encode Emotions During Speech and Song

Steven R. Livingstone and Caroline Palmer
McGill University

When speaking or singing, vocalists often move their heads in an expressive fashion, yet the influence of emotion on vocalists' head motion is unknown. Using a comparative speech/song task, we examined whether vocalists' intended emotions influence head movements and whether those movements influence the perceived emotion. In Experiment 1, vocalists were recorded with motion capture while speaking and singing each statement with different emotional intentions (very happy, happy, neutral, sad, very sad). Functional data analyses showed that head movements differed in translational and rotational displacement across emotional intentions, yet were similar across speech and song, transcending differences in F_0 (varied freely in speech, fixed in song) and lexical variability. Head motion specific to emotional state occurred before and after vocalizations, as well as during sound production, confirming that some aspects of movement were not simply a by-product of sound production. In Experiment 2, observers accurately identified vocalists' intended emotion on the basis of silent, face-occluded videos of head movements during speech and song. These results provide the first evidence that head movements encode a vocalist's emotional intent and that observers decode emotional information from these movements. We discuss implications for models of head motion during vocalizations and applied outcomes in social robotics and automated emotion recognition.

Keywords: head motion, emotion, vocal communication, speech, song

Supplemental materials: http://dx.doi.org/10.1037/emo0000106.supp

The great oratory and singing performers of our time are remarkable for their expressive and often dramatic use of head and facial movements. Are these movements superfluous, or do they serve a communicative function? Visual gestures, such as those used by speakers, musicians, and dancers, complement the auditory signal by communicating through motion. These gestures facilitate the perception of lexical and musical information, as well as expressive content (Dittrich, Troscianko, Lea, & Morgan, 1996; Ghazanfar & Takahashi, 2014; Grant & Seitz, 2000; Jesse & Massaro, 2010; Ladefoged & Johnson, 2010; McGurk & MacDonald, 1976; Munhall, Jones, Callan, Kuratate, & Vatikiotis-Bateson, 2004). During performance, musicians' visual gestures convey emotional meaning, sometimes more accurately than do their sounded performances (Carlo & Guaitella, 2004; Davidson, 1993; Vines, Krumhansl, Wanderley, Dalca, & Levitin, 2011). Whole body postures have been shown to effectively communicate expressive information to observers (Atkinson, Dittrich, Gemmell, & Young, 2004; Van den Stock, Righart, & De Gelder, 2007). To our knowledge, there has been no examination of the role of head movements in conveying emotion during vocal communication. In this article, we examine vocalists' head motion during the production of emotional speech and song, and observers' perception of emotion from these movements.

Head motion during vocalization is thought to reflect motor movements tied to the control of fundamental frequency (F_0; Munhall et al., 2004; Thompson & Russo, 2007; Thompson, Russo, & Livingstone, 2010; Yehia, Kuratate, & Vatikiotis-Bateson, 2002; Yehia, Rubin, & Vatikiotis-Bateson, 1998). Vertical raising of the head and ceiling-directional rotation of the head are both correlated with the production of a higher F_0 in speech and the size of an ascending musical interval in song (Thompson & Russo, 2007; Yehia et al., 2002, 1998). Visual presentations of vocalists' head movements have been shown to facilitate observers' perception of lexical information in speech and pitch interval size in song (Munhall et al., 2004; Thompson et al., 2010). In speech and music performance, the manipulation of F_0, or pitch, is one of the primary cues to emotional expression (Gabrielsson & Lindström, 2001; Juslin, 2001; Juslin & Vastfjall, 2008; Livingstone, Muhlberger, Brown, & Thompson, 2010; Scherer, 2003). An elevated F_0 is typically associated with higher arousal emotions, including happiness, anger, and fear, whereas a lower F_0 is associated with lower arousal emotions, including sadness and tenderness (Cowie et al., 2001; Juslin & Laukka, 2003; Scherer, 1995, 2003). Thus, vocalists' head movements may vary with the expressed emotion during speech and song due to emotion-dependent changes in F_0, a hypothesis we test here.

Steven R. Livingstone and Caroline Palmer, Department of Psychology, McGill University, Montreal, Quebec, Canada.

This research was funded in part by an ACN-Create NSERC Fellowship to Steven R. Livingstone and by a Canada Research Chair and NSERC Grant 298173 to Caroline Palmer. The authors thank Erik Koopmans, Sasha Ilnyckyj, Marcelo Wanderley, David Ostry, Frank A. Russo, Pascal van Lieshout, Frances Spidle, Rachel Brown, and Pascale Lidji for their assistance and comments.

Correspondence concerning this article should be addressed to Steven R. Livingstone, who is now at the Department of Psychology, Neuroscience & Behaviour, 1280 Main Street West, Hamilton Ontario L8S 4K1, Canada, or to Caroline Palmer, Department of Psychology, McGill University, 1205 Dr Penfield Ave, Montreal QC H3A 1B1, Canada. E-mail: steven.livingstone@mcmaster.ca or caroline.palmer@mcgill.ca

Head motion during vocalization may also arise directly from intentions to communicate specific emotional states. Darwin (1872/1965) proposed that head and body postures were representative of particular emotional states, in which the head is held upright during happiness or joy, and the head hangs down during sadness or despair (pp. 176, 210). The English language has many idioms that relate head posture to an emotional state or event. One may have his or her "head held high" during moments of happiness or pride, or be told to "keep your chin up" or to "keep your head down" during emotionally painful or dangerous situations. One may also be described as being "downcast" during periods of sadness or despair. Empirical research has corroborated these relationships to some extent, both in the encoding of expressions through head orientation and the decoding of these gestures by observers (Atkinson et al., 2004; Boone & Cunningham, 1998; Keltner, 1995; Otta, Lira, Delevati, Cesar, & Pires, 1994; Schouwstra & Hoogstraten, 1995; Tracy & Matsumoto, 2008; Tracy & Robins, 2004; Wallbott, 1998; Wallbott & Scherer, 1986). Head movements during emotional vocalization may then arise from two distinct sources—movements arising from emotion-related changes in F_0 of the voice and movements related to the visual expression of emotion through the head.

We examined vocalists' head movements during matched comparisons of emotional speech and song. During singing, F_0 is constrained by the intended musical pitch, whereas during speech, F_0 is not explicitly prescribed for nontonal languages such as English. This distinction allows one to examine the effect of emotion on head movements during speech and song while controlling for the effects of F_0 on head motion in an ecologically valid context. Vocal intensity (correlated with loudness) has also been linked to head motion in speech and song and, like F_0, varies with the expressed emotion. A louder vocal intensity is typically associated with higher arousal emotions such as happiness and anger, and a softer intensity with lower arousal emotions such as sadness and tenderness. However, the strength of the association between vocal intensity and head motion is consistently weaker than that between F_0 and head motion (Munhall et al., 2004; Thompson et al., 2010; Yehia et al., 2002) Therefore, we focus on the effect of F_0 on head movements in the present studies.

The dynamic characteristics of head movement may encode information about the emotional intentions of the vocalist. Speakers' head movements change continuously with the unfolding acoustic signal (Munhall et al., 2004; Yehia et al., 2002), and there is compelling evidence that dynamic facial information improves observers' recognition in emotional and nonemotional tasks (Ambadar, Cohn, & Reed, 2009; Atkinson et al., 2004; Bassili, 1978, 1979; Bugental, 1986; Cunningham & Wallraven, 2009; Kamachi et al., 2001; Krumhuber & Kappas, 2005; Nelson & Russell, 2014; O'Toole, Roark, & Abdi, 2002). Recent findings have suggested also that the velocity of facial expressions during expressive displays conveys important emotional information (Krumhuber, Kappas, & Manstead, 2013). Additional research has suggested that vocalists begin to form emotional facial expressions several seconds prior to vocal onset and sustain those expressions for several seconds following the offset of sound (Livingstone, Thompson, & Russo, 2009; Livingstone, Thompson, Wanderley, & Palmer, 2015; Quinto, Thompson, Kroos, & Palmer, 2014). We therefore hypothesize that head movement may encode emotion during the time periods prior to and following, as well as during, vocalization.

We used functional data analysis to examine the dynamics of head movement (Ramsay & Silverman, 2005). Functional data analysis is suited to the analysis of time-series data, such as that produced by motion capture, because it models discrete measurements as continuous mathematical functions (Ramsay & Silverman, 2002). An advantage of this approach is that events of varying duration can be compared via a process referred to as time warping, without resulting in smearing or a loss of detail (see also Livingstone, Palmer, & Schubert, 2012; Wanderley, Vines, Middleton, McKay, & Hatch, 2005). Functional analyses of variance (fANOVA) can then be used to determine time regions where significant differences occur. This is a powerful technique when used with high data-sampling rates, because it allows for precise temporal estimates of group differences. We used fANOVAs in a repeated-measures context to determine (a) whether emotion has an effect on head movements, through the use of different instructions for vocalized emotions; (b) whether lexical variability affects head movements, through the use of the same four distinct sentences in speech and song; and (c) whether vocal channel has an effect on head movements, through the comparison of speech and song.

We also addressed whether head movements associated with a vocalist's emotional intent are understood by observers. A popular theory of emotion is that affective information is conveyed using a socially shared "code" (Brunswik, 1956; Ekman & Friesen, 1969; Elfenbein & Ambady, 2002; Juslin & Laukka, 2003; Scherer, 2003). In this model, expressive information is encoded through variations in vocal, bodily, and facial features. These variations, or "cues," are associated with particular emotional states. This information is transmitted by visual, auditory, or both means, and is received and decoded by a distant subject, and the emotional information is recognized. If head movement forms part of this socially shared emotional code, then evidence of this code should be found at both production and perception (Juslin, 2000; Wiener, Devoe, Rubinow, & Geller, 1972). Specifically, we hypothesized that observers should be able to recognize a vocalist's emotional intent solely from the presence of head movements that are associated with specific emotional states.

We tested these hypotheses in two studies of production and perception of head movement during speech and song. Vocalists produced the same neutral-content sentences with instructions to emote with happy, sad, and neutral states. These emotions were selected because they are clearly communicated through the face and voice in speech and song (Juslin & Laukka, 2003; Kohler et al., 2004; Scherer, 2003). In addition, these categories exhibit low rates of confusion, suggesting they are encoded using distinct patterns of cues (Bänziger, Mortillaro, & Scherer, 2012; Juslin & Laukka, 2001). We hypothesized that these emotions may therefore elicit distinct head movement patterns. In Experiment 1, vocalists' head movements were recorded while they spoke and sang with different intended emotions. We hypothesized that vocalists' head movements would uniquely distinguish their emotional intent, while presenting similarly across speech and song. In Experiment 2, observers identified the intended emotion of vocalists from silent, face-occluded movies of head movement during speech and song. We hypothesized that observers would accurately identify the emotions of vocalists from their head movements alone.

Experiment 1

In Experiment 1, vocalists' head motion was recorded with optical motion capture while they spoke and sang four statements, each with five emotional intentions (very happy, happy, neutral, sad, and very sad). In the speaking condition, vocalists memorized the statements and produced them when cued. In the singing condition, vocalists sang the same statements with a seven-note melody that maintained the same frequencies (note pitches) across emotion conditions. On the basis of previous findings that visual cues for the head pose affect perceived emotion (in the absence of vocalization; Atkinson et al., 2004; Boone & Cunningham, 1998), we hypothesized that happy displays would exhibit upward turning of the head and that sad displays would show downward turning of the head. Previous research has also found that the magnitude of facial feature movements is correlated with the emotional intensity of the experienced event (Cacioppo, Petty, Losch, & Kim, 1986). We therefore predicted that the magnitude of vocalists' head movement would express happy and sad emotions with two levels of intensity, with more-intense emotions (very happy, very sad) exhibiting larger movements than their less-intense counterparts (happy, sad). Finally, we expected that F_0 in the speech condition would change with the intended emotion, increasing for happiness and decreasing for sadness, but would not display emotion-related changes in song (in which the pitch is prescribed). Finally, head motion was expected to change with emotional intent similarly in speech and song.

To ensure that vocalists were expressing the emotions as instructed, the stimulus recordings collected in Experiment 1 were assessed in a separate perceptual task. Participants were asked to identify vocalists' emotions from a pseudorandomly selected subset of the audio-only recordings. We chose to validate the audio recordings, because this dependent measure was orthogonal to the primary measure of study, the vocalists' head movements.

Method

Participants. Twelve adult singers (six females and six males, mean age = 22.8 years, *SD* = 6.7) were recruited from the McGill University community through postings on campus and online. Participants were native English speakers and had at least 6 years of vocal singing experience in a group (chorus, band, or other vocal ensemble; *M* = 9.83 years, *SD* = 3.0). They had varied amounts of private vocal instruction (*M* = 6.0 years, *SD* = 5.0), private musical instrument instruction (*M* = 9.0 years, *SD* = 6.7), and drama experience (*M* = 4.17 years, *SD* = 4.6). Participants were not professional actors and had only general acting experience.

Participants were screened prior to testing to ensure they had not received prior training on how to move or hold the head or face while singing; four participants were excluded on this basis. A minimum amount of group (ensemble) singing experience was used to ensure that all participants could complete the singing task without performance anxiety related to correct pitch reproduction. An inability to accurately sing the notated pitches may have elicited feelings of shame or guilt, which are known to affect head motion (Keltner, 1995; Tracy & Matsumoto, 2008). Research has indicated that vocalists are more pitch-accurate when singing with others, rather than singing alone, as seen during private singing lessons (Green, 1994).

Twenty additional adult participants (15 females and five males, mean age = 19.8 years, *SD* = 0.4) with no specific singing background were recruited for the perceptual task. These participants did not participate in any other part of the study. Participants had varied amounts of vocal experience (*M* = 2.0 years, *SD* = 0.5), private music instruction (*M* = 1.4 years, *SD* = 0.7), and drama experience (*M* = 3.2 years, *SD* = 1.0). The perceptual task took approximately 40 min, and participants received course credit for their participation.

Stimuli and apparatus. Four neutral English statements were used ("People going to the bank," "Children tapping to the beat," "Children jumping for the ball," and "People talking by the door"). Statements were seven syllables in length and matched in word frequency and familiarity using the MRC Psycholinguistic Database (Coltheart, 1981). In the song condition, a single isochronous melody (F4, F4, G4, G4, E4, E4, F4; piano MIDI tones) consisting of six eighth notes (300 ms) and ending with a quarter note (600 ms) was sounded. The melody did not contain the 3rd scale degree and was designed to be ambiguous in terms of major or minor modes, which are often associated with happy and sad emotions, respectively (Dalla Bella, Peretz, Rousseau, & Gosselin, 2001; Kastner & Crowder, 1990).

The written sentences were presented visually on a 15-in. MacBook Pro, and the melody was sounded over Sennheiser HD 500 headphones, controlled by custom Matlab (2013) software and the Psychophysics Toolbox (Brainard, 1997). Temporal accuracy of the presentation software was confirmed with the Black Box Toolkit (Plant, Hammond, & Whitehouse, 2002). An active motion capture system (NDI Optotrak Certus; spatial accuracy .1 mm) monitored the head movements of participants at a frame rate of 250 Hz. IRED markers 7 mm in diameter were placed on the participant's headphones (headband, and left and right earcups), providing a rigid body with which to align the motion coordinate system. Additional markers were also placed on the participant's face (not reported here). Vocal utterances were captured with an AKG C414 B-XLS cardioid microphone, placed 1.5 m in front of the vocalists, at 44 kHz. Sound recordings were synchronized with motion data via the Optotrak Data Acquisition Unit.

Ten acoustic recordings were selected from each vocalist's performances in the production task for use in the perceptual task. One vocalization for each of the five emotions was randomly selected from both the speech and song conditions. A subset of 120 recordings was chosen from the full set of 960 trials because it balanced the needs of establishing confidence in the stimuli with having an appropriate experimental duration for a repeated-measures perceptual study. Trials for each vocalist–channel–emotion condition were numbered, and a custom Matlab script incorporating the Mersenne Twister random number generator was used to identify the recordings for inclusion. Recordings were then trimmed using Adobe Audition to begin 1 s prior to vocal onset and to end 1 s after vocal offset. Stimuli were presented auditorily over Sennheiser HD 500 headphones, on a 21-in. iMac controlled by custom Matlab software and the Psychophysics Toolbox (Brainard, 1997).

To confirm that trials selected for the perceptual task were representative of the full sample, we examined acoustic measures including mean F_0, vocal intensity, and duration (see Analyses: Vocal acoustics for a description of acoustic analyses). These acoustic measures were selected because they are primary cues to vocal emotion (Cowie et al., 2001; Scherer, 2003). Mean values of the seven unselected vocalist–channel–emotion trials were compared with those of the selected trials using separate paired-samples *t* tests. The perceptual task stimuli were not significantly different from the full sample on any of the acoustic measures: F_0, $t(119) = 1.23$, $p = .22$; vocal intensity, $t(119) = 0.2$, $p = .85$; and duration, $t(119) = 0.27$, $p = .79$. These results provide additional confidence that the trials selected for the perceptual task were representative of the full sample.

Design and procedure.

Production task. The experimental design was a Vocal Channel (2 levels: speech or song) × Emotion (5 levels: neutral, happy, very happy, sad, very sad) × Statement (4 statements mentioned earlier in the Stimuli and Apparatus section) × Repetition (2 levels) within-subject design, with 80 trials per participant. Trials were blocked by vocal channel, with speech presented first to avoid any influence of pitch or timing from the regular pace of the song condition. Trials were blocked by statement and emotion (counterbalanced) to reduce fatigue.

Participants were told to prepare themselves emotionally as they would for a live performance and were given time between blocks to prepare themselves. Vocalists were given no instruction regarding their composure leading up to or following the offset of vocal sound and were told only to speak or sing in "an expressive manner as though performing to an audience." This procedure was used to induce the mental and physiological correlates of emotion that vocalists typically display and experience when performing in front of others. The use of induction techniques is gaining popularity among researchers who seek ecologically valid recordings of emotion in a laboratory setting (Bänziger et al., 2012; Douglas-Cowie et al., 2007; Kaulard, Cunningham, Bülthoff, & Wallraven, 2012; Livingstone, Choi, & Russo, 2014).

Participants began with a series of speech practice trials; the practice statements differed from those presented in the experimental trials. On each trial, participants read and memorized the statement presented on a computer screen and were told to wait for a green light to appear before they began their vocalization. The trial time line, presented in Figure 1, consisted of four main epochs: stimulus presentation (visually displayed statement), countdown timer (4–3–2–1), begin vocalization (green light), and end of vocalization. Practice trials were repeated until participants were comfortable with the task. In the experimental block, participants were first shown the four statements that would be used throughout the block. They then completed the speech experimental trials. After a rest break, they completed a series of song practice trials (with the same statements as those in the speech practice trials) in which they read and memorized the statement presented on a computer screen and heard a short melody. Participants were told to sing one syllable per tone, using the pitches and timing of the presented melody—for example, peo(1, F4)-ple(2, F4) talk(3, G4)-ing(4, G4) by(5, E4) the(6, E4) door(7, F4). In the experimental block, participants were shown a new statement on each trial, after which the melody was presented; they were told to begin singing when the green light appeared (following the timing of Figure 1). Trials were repeated if participants made a mistake or if they moved outside the motion capture volume.

Perceptual task. The experimental design for the perceptual task was a Vocalist (12 levels) × Vocal Channel (2 levels: speech or song) × Emotion (5 levels: neutral, happy, very happy, sad, very sad) × Repetition (2 levels) within-subject design, with 240 trials per participant. Trials were blocked by vocal channel, with vocalist,

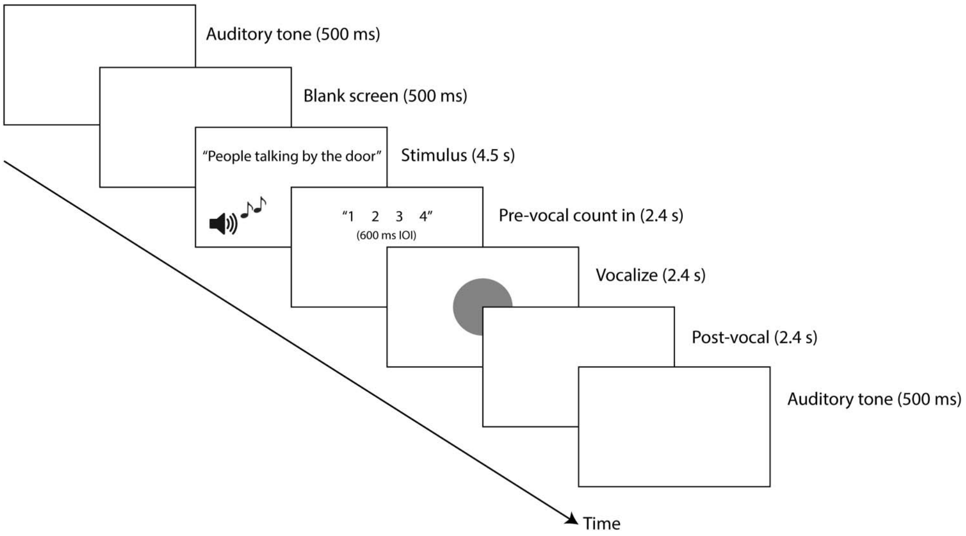

Figure 1. Time line of trials in Experiment 1. Each trial began with a 500-ms auditory tone, followed by 500 ms of blank screen. The statement to be spoken or sung was then presented. In the song condition, the melody was also sounded. A prevocal count-in timer was then presented. Participants began vocalization with the appearance of the green circle. Additional movements were captured during the postvocal epoch (blank screen). The trial ended with a 500-ms auditory tone. Head motion and acoustic information were captured throughout the entire trial time line.

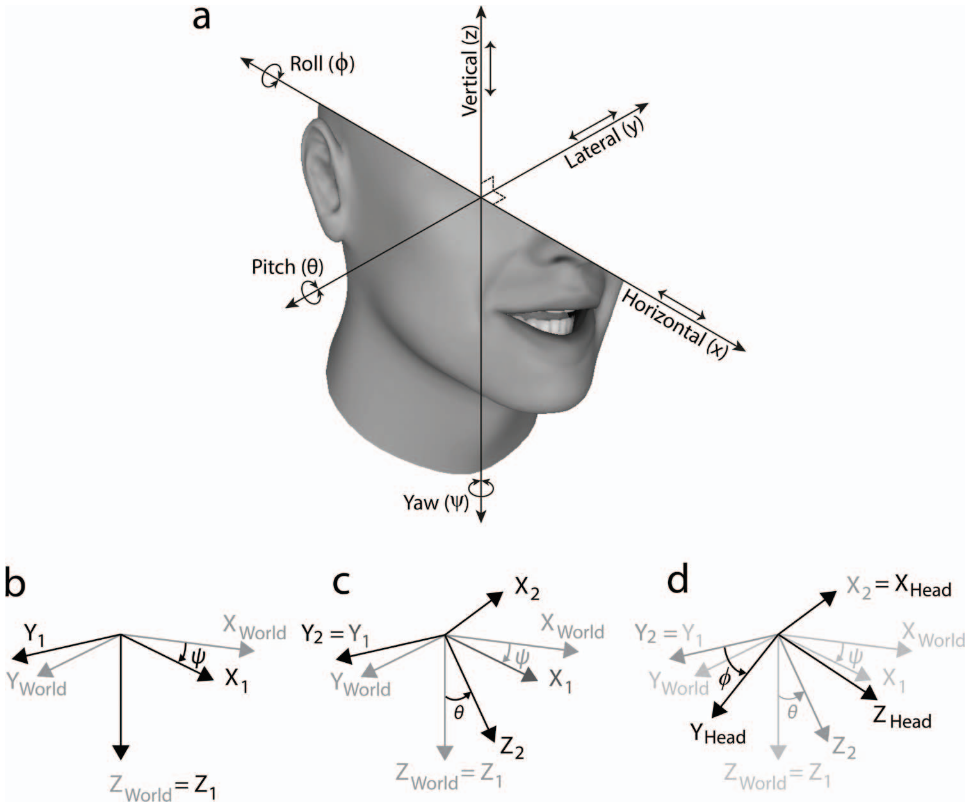

Figure 2. (a) The frame of reference of a participant's head. The three axes illustrate the three directions of translational motion—*horizontal* (forward–backward, *x*), *lateral* (left–right, *y*), and *vertical* (up–down, *z*)—and the three angles of rotational movement—*yaw* (looking left–right, ψ), *pitch* (looking up–down, θ), and *roll* (ear to either shoulder, ϕ). The origin is formed by the lateral axis spanned by the bilateral tragus. Lower panes demonstrate the three successive rotation operations that transform data to the local coordinate system of the participant's head, showing (b) rotation angle ψ around z_{World}, $\mathbf{R}_z = (\psi)$; (c) rotation angle θ around y_1, $\mathbf{R}_y = (\theta)$; and (d) rotation angle ϕ around x_2, $\mathbf{R}_x = (\phi)$.

emotion, and repetition presented in random order. On each trial, participants were asked to identify the emotion of the vocalist using a forced-choice categorical response measure (neutral, happy, sad). Participants also completed a background questionnaire about their singing, musical, and drama experience.

Head motion processing. Prior to statistical analysis, head motion data were transformed from the fixed world coordinate system of the motion capture system to the moving time-dependent coordinate system of the participant's head. This transformation enabled the examination of the full range of translational and rotational head movements, referred to as six degrees of freedom of motion (6DoF). This approach represents an important methodological improvement over "point-to-point" Euclidean distance analysis (1DoF), which reflects the magnitude but not the direction or rotation of movement. This improvement is critical to the study of directions of movement that differentiate emotions (e.g., looking upward and raising lip corners for happiness vs. looking downward and depressing lip corners for sadness), because a mathematical exposition of coordinate system rotation can facilitate future research. Custom Matlab scripts were used to perform the rotation analysis, rather than the default NDI rigid body algorithms.

A three-marker rigid body formed by the axes of the participant's headphones was used to align the local coordinate system of the participant's head, as illustrated in Figure 2a.

At each time point, motion capture data were transformed from the fixed world coordinate system (w) to the moving head coordinate system (h), as defined by Equation 1:

$$\begin{pmatrix} h_x \\ h_y \\ h_z \end{pmatrix} = R_{\mathrm{Head}}^{\mathrm{World}} \cdot \begin{pmatrix} w_x \\ w_y \\ w_z \end{pmatrix}. \tag{1}$$

The transformation matrix $R_{\mathrm{Head}}^{\mathrm{World}}$ at each time point was calculated by three Tait-Bryan intrinsic axial rotations (z–y–x), illustrated in Figure 2b–d, respectively, as defined by Equation 2:

$$R_{\mathrm{Head}}^{\mathrm{World}} = R_z(\psi) \cdot R_y(\theta) \cdot R_x(\phi) = \begin{pmatrix} r_{xx} & r_{xy} & r_{xz} \\ r_{yx} & r_{yy} & r_{yz} \\ r_{zx} & r_{zy} & r_{zz} \end{pmatrix}. \tag{2}$$

Rotation around the z, y, and x axes were defined, respectively, in Equation 3:

$$R_z(\psi) = \begin{pmatrix} \cos\psi & -\sin\psi & 0 \\ \sin\psi & \cos\psi & 0 \\ 0 & 0 & 1 \end{pmatrix},\ R_y(\theta) = \begin{pmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{pmatrix},\ R_x(\phi) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\phi & -\sin\phi \\ 0 & \sin\phi & \cos\phi \end{pmatrix}. \tag{3}$$

When factorized, the final rotation operation $R^{\text{World}}_{\text{Head}}$ was defined by Equation 4:

$$R^{\text{World}}_{\text{Head}} = \begin{pmatrix} \cos\psi\cos\theta & -\sin\psi\cos\phi + \cos\psi\sin\theta\sin\phi & \sin\psi\sin\phi + \cos\psi\sin\theta\cos\phi \\ \sin\psi\cos\theta & \cos\psi\cos\phi + \sin\psi\sin\theta\sin\phi & -\cos\psi\ \sin\phi + \sin\psi\sin\theta\cos\phi \\ -\sin\theta & \cos\theta\sin\phi & \cos\theta\cos\phi \end{pmatrix}. \tag{4}$$

Head rotation angles (in radians) were extracted from the rotated data: *yaw* (ψ) rotation around the *z* axis, which corresponded to turning the head left or right; *pitch* (θ) rotation around the *y* axis, which corresponded to tilting the head up to the ceiling or down to the floor; and *roll* (ϕ) rotation around the *x* axis, which corresponded to leaning the head to either shoulder. These measures are defined, respectively, in Equation 5:

$$\text{yaw}(\psi) = \tan^{-1}\left(r_{yx}/r_{xx}\right),\ \text{pitch}(\theta) = \sin^{-1}(r_{zx}),\ \text{roll}(\phi) = \tan^{-1}\left(r_{zy}/r_{zz}\right). \tag{5}$$

Translational motion of the participant's head (in mm), representing the three axes of motion were defined as *horizontal* (*x*) movement along the *x* axis, which corresponds to moving forward or backward; *lateral* (*y*) movement along the *y* axis, which corresponds to moving left or right; and *vertical* (*z*) movement along the z axis, which corresponds to moving up or down.

Analyses.

Head movements. Motion data were set to zero for a "neutral resting" position of the participant's head during a 2,000-ms baseline window prior to each trial onset, using a baseline subtraction procedure. The modal value within this baseline window was subtracted from displacement values during the trial time line. The baseline-adjusted marker trajectories represented how the head position deviated throughout the trial from its resting position.

Functional data analyses were performed using the Functional Data Analysis toolbox (Ramsay & Silverman, 2010). Occasional missing data were interpolated (less than 0.0001% of data). Order 6 B-splines were fit to the second derivative of marker trajectories with a ratio of 1:4 knots to data samples. The data were smoothed using a roughness penalty on the fourth derivative ($\lambda = 10^{-8}$). Feature registration (time warping) was used to align trajectories to the syllable boundaries (six events) as determined from the acoustic analyses, described in Analyses: Vocal acoustics. Functional data were resampled to produce 75 equally spaced data points per syllable (300 ms at 250 Hz) for the first six syllables, with the final syllable resampled to 150 data points (600 ms at 250 Hz). This derivation enabled a syllable-matched comparison of head movements across speech and song. Functional analyses of variance (fANOVA) were used to examine whether differences were present in the motion trajectories at each time point (see also Livingstone et al., 2012). Significance levels were corrected for multiple comparisons with False discovery rate using the Benjamini–Hochberg–Yekutieli procedure for dependent statistical tests, with a *q* value of .05 (Benjamini & Hochberg, 1995; Benjamini & Yekutieli, 2001). We report mean *F* statistic, *p* value, and eta-squared values across time regions that reached statistical significance ($p < .05$). Motion effect sizes are reported as eta-squared values.

Vocal acoustics. Acoustic recordings were analyzed with Praat (Boersma & Weenink, 2010). Speech utterances were segmented at syllable boundaries and were coded by a rater; 8% of the samples were checked by a second rater (mean interrater boundary time difference = 0.0026 s, *SD* = 0.0024 s). Boundaries were determined by changes in the spectrogram, and in the F_0 and acoustic intensity contours. Singing trials were inspected by the raters to confirm that vocalists had sung the prescribed melody. F_0 contours were extracted using an autocorrelation algorithm (ac) in Praat. Mean F_0 values across the entire utterance were analyzed with a mixed-design analysis of variance (ANOVA). When Mauchly's sphericity test was significant, Greenhouse-Geisser's correction was applied. Acoustic effect sizes report partial eta-squared values. All reported means are accompanied by 95% confidence intervals in square brackets. Pairwise comparisons were adjusted using Bonferonni correction. All statistical tests were conducted in Matlab 2013b and SPSS 20.0.0 (IBM Corp, 2011).

Perceptual task. Data from the perceptual task were coded as 1 (*correct*) when the emotional category selected by the listener matched the emotional category that the vocalist had been instructed to express and 2 (*incorrect*) otherwise. We use the term *accuracy* to refer to the proportion of correct responses (see also Bänziger et al., 2012; Juslin & Laukka, 2001). Production trials containing happy and very happy instructions were coded as correct when *happy* was selected, trials containing sad and very sad instructions were coded as correct if *sad* was selected, and trials containing neutral instructions were coded as correct if *neutral* was selected. Mean accuracy scores are reported to ease readability and to facilitate comparison with existing perceptual studies of vocal emotion. One-sample *t* tests were conducted for each vocal–channel–emotion category to determine whether accuracy ratings differed from chance. Significance levels of *t* tests were corrected for multiple comparisons with False discovery rate and a *q* value of .05.

Results

Head movements. Because this was the first study to report head movements in a dynamic vocal context we analyzed the full six degrees of freedom (6DoF) of head movements, which included horizontal, vertical, and lateral displacement and yaw, pitch, and roll rotation angles (see Atkinson et al., 2004; Keltner, 1995, for an examination of nonvocal head movement during emotional expression).

A three-way repeated-measures fANOVA was conducted on vocalists' horizontal head measures (deviation from resting position of head) by channel (2 levels: speech, song), emotion (5 levels: very happy, happy, neutral, sad, very sad), and statement (4 statements mentioned in the Stimuli and Apparatus section). No effect of channel or its interactions was found in the analysis, confirming that speech and song exhibited similar horizontal head movements. Figure 3a shows the mean horizontal head displacement (forward-backward movement, *x*) values across all trials by emotion conditions. Regions of statistical significance ($p < .05$) are indicated by the black horizontal bar in the time line. The main effect of emotion was found during the majority of vocal sound productions, $F(4, 44) = 7.17$, $p < .01$, $\eta^2 = .05$. Happy and sad utterances were characterized by larger forward movement of the head than were emotionally neutral utterances, with very happy utterances appearing to move forward quickly at the onset of vocalization. Very happy and very sad productions also appeared to exhibit larger forward displacement relative to their happy and sad counterparts, respectively. No effect of statement or its interactions was found in the analysis, confirming that lexical variability did not affect horizontal head movements.

The previous analysis suggested that the velocity of horizontal head movements may vary with the expressed emotion. To examine this relationship, we conducted a follow-up analysis that involved a three-way repeated-measures fANOVA on the velocity of vocalists' horizontal head movements (forward–backward, $\dot{x}$). No effect of channel or its interactions was found in the analysis, confirming that speech and song exhibited similar velocities of horizontal movements. Figure 3b shows the mean velocity of vocalists' horizontal head movements across all trials by emotion conditions. A main effect of emotion was found, $F(4, 44) = 9.26$, $p < .01$, $\eta^2 = .13$, which began 100 ms prior to vocal onset and continued for 220 ms during the start of vocal sound production. These results suggest that the forward-moving velocity of head movements distinguished the vocalized emotion and that very happy utterances exhibited greater forward-moving velocity than did other emotions. No effect of statement or its interactions was found in the analysis, confirming that lexical variability did not affect the velocity of horizontal head movements.

A three-way repeated-measures fANOVA was conducted on vocalists' vertical head measures (moving up–down, *z*). No effect of channel or its interactions was found in the analysis, confirming that speech and song exhibited similar vertical movements. Figure 3c shows vocalists' mean vertical head displacement (deviation from resting position) across all trials by emotion conditions. A

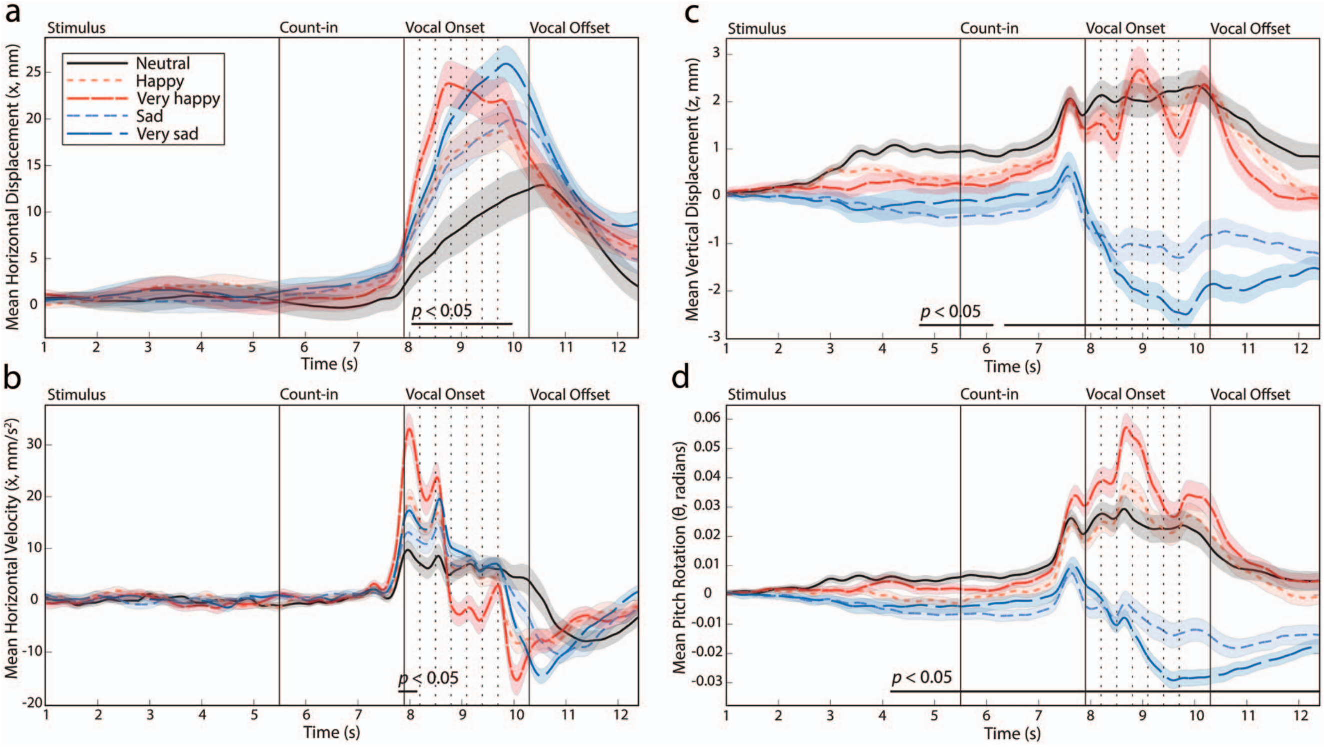

Figure 3. Main effects of emotion conditions on four aspects of head movements in Experiment 1. Each trajectory line is the functionally time-warped mean across all actors, vocal channels, statements, and repetitions (192 trials per trajectory line). Zero represents the neutral "at rest" position of the head. Dashed vertical lines between vocal onset and offset indicate syllable boundaries. Black horizontal lines below trajectories indicate regions of significance at $p < .05$. Error bars are indicated by shaded regions around trajectory lines. (a) Mean horizontal displacement of head movement (forward–backward, *x*). (b) Mean horizontal velocity of head movement (forward–backward, $\dot{X}$). (c) Mean vertical displacement of head (up–down, *z*). (d) Mean rotational pitch of head movement (looking up–down, θ).

significant main effect of emotion was found throughout vocalization, which continued long after vocal sound ended, $F(4, 44) = 8.47$, $p < .01$, $\eta^2 = .11$. Importantly, these effects began 3,196 ms prior to the onset of vocal sound, confirming that vocalists' head movements are not only a result of vocal sound production. These results indicated that vocalists vertically raised their head during happy and neutral vocalizations and lowered their head during sad vocalizations. Very sad productions appeared to exhibit larger downward vertical displacement relative to sad productions. A similar relationship was not found for happy vocalizations. No effect of statement or its interactions was found in the analysis, confirming that lexical content did not affect vertical head movements. We then analyzed lateral (left–right, y) head displacement with a three-way repeated-measures fANOVA. No significant effects were indicated. The speed of vertical head movements did not vary noticeably with the expressed emotion, and so vertical velocity was not examined.

We next examined vocalists' rotational head movements. A three-way repeated-measures fANOVA was conducted on vocalists' rotational pitch (tilting up–down, θ) head movements. No effect of channel or its interactions was found in the analysis, confirming that speech and song exhibited similar rotational pitch movements. Figure 3d shows vocalists' mean rotational pitch values across all trials by emotion conditions. Regions of statistical significance ($p < .05$) are indicated by the black horizontal bar in the time line. A significant main effect of emotion was found throughout vocalization, which continued after vocal sound ended, $F(4, 44) = 8.93$, $p = .005$, $\eta^2 = .10$. Importantly, these effects began 3,748 ms prior to the onset of vocal sound, beginning during the task presentation epoch. These results corroborate the findings for vertical head movements, confirming that vocalists' head movements are not simply a result of vocal sound production. Vocalists appeared to tilt their heads upward for happy utterances and downward for sad utterances. Very happy and very sad productions also appeared to exhibit larger rotational pitch relative to their happy and sad counterparts, respectively, with very happy showing more ceiling-directed rotation relative to happy, and very sad showing more floor-looking rotation relative to sadness. No effect of statement or its interactions was found in the analysis. The speed of pitch rotation movements did not appear to vary noticeably with the expressed emotion, and so pitch velocity was not examined. We then analyzed yaw (turning left–right, ψ) and roll (ear to either shoulder, ϕ) measures. No significant effects were indicated for either rotational angle.

Collectively, these results suggest that vocalists' head movements varied consistently with the vocalized emotion. Importantly, these movement trajectories were comparable across vocal channels (speech and song) and lexical variability (four different statements). Furthermore, emotionally dependent movements began well before the onset of vocal sound for vertical and rotational pitch movements, confirming that some aspects of head movement are not simply a by-product of vocal sound production. These results provide preliminary support for the hypothesis that head movements function to visually express emotion.

Voice acoustics. We next examined the prediction that F_0 would vary characteristically with emotion in speech but would remain constant across emotional conditions in song. Mean F_0 values of speech utterances were analyzed in a three-way mixed-design ANOVA, with emotion (5) and statement (4) entered as within-subject factors and gender (2) entered as a between-subjects factor. A main effect of emotion was found, $F(1.74, 17.40) = 15.53$, $p < .001$, $\eta_p^2 = .61$. Pairwise comparisons confirmed that very happy speech (M = 253.34 Hz, 95% CI [215.47, 291.2]) yielded a higher F_0 than did happy (M = 211.4 Hz, [169.73, 253.08]), neutral (M = 175.57 Hz, [143.38, 207.76]), sad (M = 186.75 Hz, [143.95, 229.55]), and very sad (M = 199.21 Hz, [159.37, 239.06]) spoken utterances. In addition, happy speech yielder a higher F_0 than did neutral speech. No effect of statement was found, $F(3, 33) = 1.94$, $p = .14$. As expected, a main effect of gender was found, $F(1, 10) = 57.72$, $p < .001$, $\eta_p^2 = .85$. No interactions with gender were indicated (all $ps > .50$). These results confirm that vocalists varied their F_0 in systematic ways across emotions in speech.

Next we analyzed mean F_0 of song utterances in a three-way mixed-design ANOVA. Importantly, no effect of emotion was found, $F(4, 21.14) = 0.11$, $p = .98$. No effect of statement was found, $F(3, 11.92) = 1.27$, $p = .30$. As expected, a main effect of gender was reported, $F(1, 10) = 11{,}704.67$, $p < .001$, $\eta_p^2 = .99$. No interactions with gender were reported (all $ps > .45$). Collectively, these results confirm the prediction that F_0 varies systematically across emotions in speech but is held relatively constant across emotions in song.

Relating head movements and vocal acoustics. To examine whether head motion was related to vocalists' F_0, we regressed the voiced portions of each F_0 time series contour on the six degrees of head motion, using separate multiple linear regressions for every trial ($n = 960$). This analysis replicates that performed by Munhall et al. (2004) and Yehia et al. (2002). The predictors were horizontal, lateral, and vertical displacement, yaw, pitch, and roll rotation angles, while the criterion variable was F_0. A three-way repeated-measures ANOVA was conducted on the resultant squared correlation coefficients, representing variance accounted for, with vocal channel (2), emotion (5), and statement (4) entered as within-subject factors. A main effect of vocal channel was reported, $F(1, 11) = 16.7$, $p = .002$, $\eta_p^2 = .6$, in which the mean percentage of variance in F_0 accounted for by motion of the head in speech (M = 76.4%, [.73, .80]) was significantly greater than the mean variance accounted for in singing (M = 67.0%, [.64, .70]). A main effect of emotion was also reported, $F(2.44, 26.89) = 4.13$, $p = .021$, $\eta_p^2 = .27$, with very happy (M = 68.2%, [.65, .71]), happy (M = 69.6%, [.66, .73]), neutral (M = 74.3%, [.70, .79]), sad (M = 75.4%, [.73, .78]), and very sad (M = 71%, [.66, .76]). Pairwise comparisons confirmed that very happy vocalizations accounted for less of the variance in F_0 than did sad vocalizations. A main effect of statement was also found, $F(4, 33) = 3.09$, $p = .04$, $\eta_p^2 = .22$, with the "Door (M = 73.4%, [.71, .76]), "Ball" (M = 69.9%, [.67, .73]), "Bank" (M = 72.3%, [.70, .76]), and "Beat" (M = 70.8%, [.68, .74]) statements. Pairwise comparisons confirmed that "Bank" utterances accounted for more of the variance than did "Ball" utterances. These results indicate the variation in F_0 explained by head movements varied to some extent with the expressed emotion and statement. No other main effects or interactions were found. These results confirm that although head motion was related to F_0, head motion became less predictive of F_0 during vocal production of a fixed melody than during speech.

As a complement to the previous analysis, a test was done to examine whether head movements prior to vocal production were related to F_0. We chose movements prior to vocal production, because movements succeeding vocalization are directly influenced by the preceding movements used in the control of F_0. We regressed the mean F_0 of an utterance on the means of the six degrees of head motion during the time period from stimulus presentation to vocal onset for that same utterance. A stepwise regression with backward elimination was performed. For the stepping method criteria, the p value for including a variable was set at .05 and the p value for excluding a variable at .10. The model with the final adjusted R^2 was chosen. Head movements in the absence of vocal sound were found to explain a significant proportion of variance in F_0, $R^2 = 0.02$, $F(3, 959) = 7.67$, $p < .001$. Horizontal, $\beta = -.099$, $t(959) = -3.09$, $p = .002$; vertical, $\beta = .088$, $t(959) = 2.74$, $p = .006$; and yaw, $\beta = -.078$, $t(959) = -2.44$, $p = .015$, movements were found to significantly predict F_0 values. However, although head movements in the 7 s preceding vocal production were related to their subsequent F_0, the movements explained a much smaller percentage of the variance than did the movements that occurred during vocal sound production (2% vs. 71.7%, respectively). This relationship is to be expected, because movements preceding vocal sound and F_0 during production both reflect the expressed emotion. These results further confirm that vocalists' F_0 are related to their head movements, but only to an extent.

Perceptual task. Mean response rates for each of the three emotion categories in the perceptual task are shown in Table 1. Participants' mean accuracy across all utterances was $M = .71$, $SE = .02$. To determine whether observers' accuracy scores differed from chance, we conducted separate one-sample t tests with a chance estimate of .33 for each vocal–channel–emotion category. Significance levels were corrected for multiple comparisons with False discovery rate and a q value of .05 (Benjamini & Hochberg, 1995). All 10 vocal channel–emotion categories were correctly identified significantly above chance (.33), indicating the stimuli were perceived on the basis of their vocal content to match the instructed emotional state. Accuracy scores and their variation between emotional categories were comparable to those commonly reported in the vocal emotion literature, where neutral, happiness, and sadness are identified at approximately 74%, 57%, and 71%, respectively (Juslin & Laukka, 2003; Scherer, 2003).

Discussion

Vocalists exhibited different patterns of head motion that varied with the intended emotion. During sad vocalizations, vocalists exhibited a "downcast" posture, with vertical lowering of the head and floor-looking rotational pitch of the head. These movement patterns differed systematically from those during happy vocalizations, which exhibited ceiling-directed rotational pitch, earlier onset, and higher velocity of forward-moving horizontal movement. Neutral vocalizations generally exhibited an attenuated pattern of movements relative to other emotions. Very happy and very sad productions generally showed a greater magnitude of head movement relative to their happy and sad counterparts. Head movement trajectories were similar across vocal channels (speech, song) and varying lexical content (four statements), transcending differences in speech and song, and lexical variability. Emotion-based head movement also occurred immediately before and after sound production, with vertical and rotational pitch movements beginning before the onset of vocal sound and continuing after vocal offset. Collectively, these results suggest that head movements reflect the emotional intent of the vocalist, and those aspects of expressive head movement are not simply a by-product of vocal sound production.

Fundamental frequency of the voice varied with emotion during speech, as suggested by previous research (Cowie et al., 2001), and remained constant across emotions during song, as predicted for the fixed-pitch melody. Furthermore, head motion was correlated with F_0, more so for speech than for song. Vocalists' F_0 in speech increased for happy and very happy utterances relative to neutral, sad, and very sad vocalizations (Cowie et al., 2001; Scherer, 2003). Consistent with previous reports of mixed findings when comparing F_0 of speech emoted with neutral and sad states (Scherer, Banse, Wallbott, & Goldbeck, 1991), vocalists' F_0 for sad and very sad utterances did not decrease relative to neutral utterances. In sum, changes in F_0 in speech coincide with head motion, but F_0 alone cannot account for observed differences in head movements during emotional vocal communication.

The perceptual task confirmed that vocalists in Experiment 1 were producing the instructed expressions of emotion. Perceivers' accuracy rates were comparable to those reported in the literature for emotional speech and singing tasks (Cowie et al., 2001; Juslin & Laukka, 2003; Livingstone et al., 2015; Scherer, 2003). The perceptual task was an important methodological check, because it provides confidence that the recordings analyzed in Experiment 1 were representative expressions of emotion.

The results of Experiment 1 confirmed that vocalists exhibited distinct patterns of head movements that varied with the emotional intent. Some aspects of expressive head movements began prior to and continued after sound production. These findings suggest that

Table 1
Mean Accuracy Scores (% Correct) for the Perceptual Task in Experiment 1 and Results of Separate One-Sample t-Tests With a Chance Estimate of .33

	Speech			Song		
Emotion	Mean	*SE*	Chance test	Mean	*SE*	Chance test
Neutral	.84	.03	$t(19) = 19.45$, $p < .001$	.74	.04	$t(19) = 10.98$, $p < .001$
Happy	.56	.04	$t(19) = 5.42$, $p < .001$	.40	.03	$t(19) = 2.45$, $p = .024$
Very Happy	.93	.03	$t(19) = 21.79$, $p < .001$	.69	.03	$t(19) = 13.78$, $p < .001$
Sad	.75	.04	$t(19) = 11.09$, $p < .001$	.64	.04	$t(19) = 7.81$, $p < .001$
Very Sad	.87	.03	$t(19) = 20.42$, $p < .001$	.70	.03	$t(19) = 12.35$, $p < .001$

Note. p values are corrected for multiple comparisons.

vocalists' head movements function as a visual form of emotional communication. We examined this hypothesis further in the next experiment.

Experiment 2

In Experiment 2, viewers identified vocalists' emotional intent from silent visual displays of their head movement during speech and song. Silent video recordings of vocalists' heads with facial features occluded were presented, such that only whole-head motion cues were present. Because similar head movements for speech and song were found in Experiment 1, we hypothesized that observers would reliably distinguish the intended emotion of the vocalists from their head motion with similar levels of accuracy for both speech and song.

Method

Participants. Twenty-four native English-speaking adults (17 female and seven male, mean age = 21.6 years, *SD* = 3.9) were recruited from the McGill community. Participants had received varied amounts of private music instruction (*M* = 5.5 years, *SD* = 5.3), singing experience (*M* = 1.7 years, *SD* = 2.5), and drama experience (*M* = 0.7 years, *SD* = 2.1). No participants from the previous experiment participated in Experiment 2.

Stimuli and apparatus. Two singers (one female, one male) who participated in Experiment 1 were asked to participate in Experiment 2. The two singers were selected because their level of vocal experience was similar to the group mean in Experiment 1, Singer 1 (*M* = 9 years); Singer 2 (*M* = 6 years); group (*M* = 9.83 years), and their vocalizations received high rates of emotion identification accuracy in the perceptual task (*M* = .79). Two additional participants from Experiment 1 with similar levels of experience were invited to participate but declined. The two participating singers were invited back for a new experimental session and were video-recorded while speaking and singing statements with the emotional intentions of very happy, neutral, and very sad. The vocalists were given no further information about the goals of the experiment. Statements, melody, and trial structure were the same as in Experiment 1. Vocalists were recorded with a JVC Everio GZ-HD6 video camera and an AKG C414 B-XLS cardioid microphone. Vocalists stood in front of a green screen cloth, illuminated with three Cameron Quartz Imager Q-750 lights with white-diffusion parabolic umbrellas.

Stimulus recordings were trimmed to contain only the time region during vocalized sound, to ensure that observers were identifying emotion on the basis of head movements that occurred during vocal sound production. Vocal time regions were determined with Praat (Boersma & Weenink, 2010), and video recordings were edited with Adobe Premiere Pro CS5. Chroma key compositing (green screen replacement) was applied, with the background set to a single green color (hex #03804e). Video recordings of vocalists' faces were then occluded using skin-toned ellipses; an example is shown in Figure 4. Compositing of head motion tracking and ellipse layer were performed with Adobe After Effects CS5 (see also the movie in the online supplemental material). Stimuli were presented with E-Prime 2.0 software (Psychology Software Tools, Pittsburgh, PA) on a windows PC, connected to a 19-in. Dell flat-screen LCD monitor, at a distance of approximately 1 m from the participant.

Design and procedure. The experiment used a Vocal Channel (2 levels: speech or song) × Vocalist (2 levels) × Emotion (3 levels: happy, neutral, sad) × Statement (4 statements mentioned in Stimuli and Apparatus section of Experiment 1) × Repetition (2 levels) within-subject design with 96 trials per participant. Trials were blocked and counterbalanced by production and presented in a pseudorandom order.

Participants sat in front of a computer and were told that they would be presented with short videos of people speaking and singing with different emotions. They were told that the videos would have no sound and that people's faces would be occluded. Participants were asked to identify the emotion they thought the vocalist was expressing, using a forced-choice categorical re-

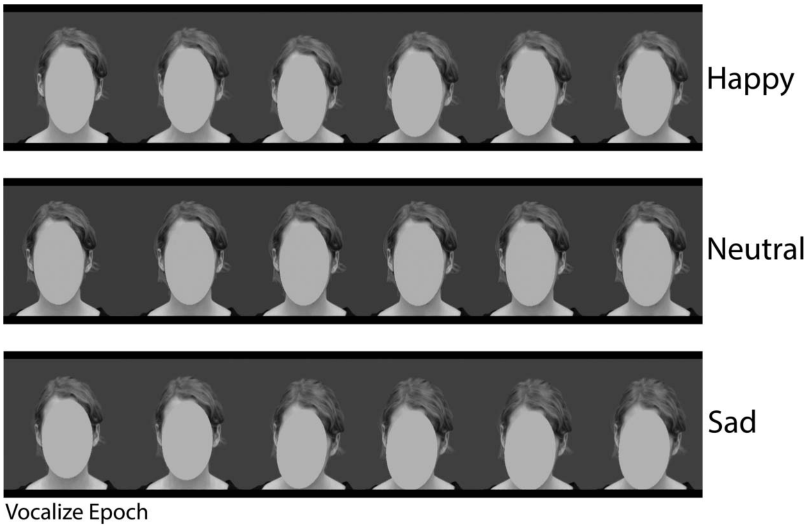

Figure 4. Still images from happy (top row), neutral (middle row), and sad (bottom row) movie stimuli used in Experiment 2. Downward-tilted rotational head pitch is visible in the sad vocalization example.

sponse measure (neutral, happy, sad). Participants also completed a background questionnaire about their singing, acting, and musical experience. Participation in the experiment took approximately 30 min.

Data were coded as 1 (*correct*) when the emotional category selected by the observer matched the emotional category that the vocalist had been asked to portray and 0 (*incorrect*) otherwise. We used the term *accuracy* to refer to the proportion of correct responses. To correct for possible response bias in the multilevel response task, we also calculated unbiased hit rates (Wagner, 1993). Unbiased hit rate scores are proportion scores (0–1) and were arcsine square root transformed prior to analysis. The factors vocalist, statement, and repetition were collapsed over during analysis. Accuracy and unbiased hit rate scores were analyzed with repeated-measures ANOVAs. When Mauchly's sphericity test was significant, Greenhouse-Geisser's correction was applied. Effect sizes are reported with partial eta-squared values, means are accompanied by 95% confidence intervals, and pairwise comparisons were adjusted using Bonferonni correction. Statistical tests were conducted in Matlab 2013b and SPSS v20.0.0.

Results

A two-way repeated-measures analysis of variance (ANOVA) was conducted on observers' mean accuracy (% correct) with vocal channel (2) and emotion (3) as within-subject factors; mean values are shown in Figure 5. A significant main effect of emotion was reported, $F(1, 23) = 26.49$, $p < .001$, $\eta_p^2 = .54$, with higher accuracy for neutral (M = 79.8%, 95% CI [.73, .86]) and happy (M = 69.8%, [.64, .76]) than for sad (M = 55.9%, [.51, .61]) emotions. In addition, a two-way interaction of Channel × Emotion was reported, $F(2, 46) = 6.57$, $p = .003$, $\eta_p^2 = .22$, illustrated in Figure 5. Post hoc comparisons (Tukey's HSD = .1, $\alpha = .05$) confirmed that the neutral emotion was identified more accurately in speech (M = 84.9%, 95% CI [.79, .91]) than in song (M = 74.7%, [.66, .84]), whereas happy and sad did not differ between channels. No significant effect of vocal channel was reported, $F(1, 23) = 3.59$, $p = .071$, with speech (M = 70.2%, [.65, .75]) and song (M = 66.8%, [.62, .71]).

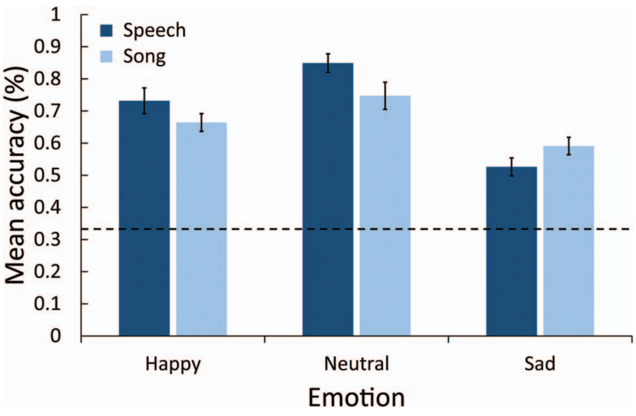

Figure 5. Observers' mean accuracy (% correct) by emotion and vocal channel in Experiment 2. Dashed horizontal line indicates chance performance (.33). Error bars represent 1 standard error of the mean. See the online article for a color version of this figure.

To determine whether observers' accuracy scores differed from chance, we conducted one-sample *t* tests with a chance estimate of .33. Observers identified the emotional intent of vocalists overall well above chance levels (M = 68.5%, [.64, .73]), $t(23) = 16.99$, $p < .001$, $d = 7.09$. Within each emotion, vocalists' intended expressions were all identified above the level of chance: happy (M = 69.8%), $t(23) = 12.77$, $p < .001$, $d = 5.32$; neutral (M = 79.8%), $t(23) = 14.74$, $p < .001$, $d = 6.15$; and sad (M = 55.9%), $t(23) = 9.63$, $p < .001$, $d = 4.02$. These results confirm that observers accurately decoded vocalists' intended emotions above chance on the basis of silent face-occluded head-motion displays.

To correct for possible response bias in multilevel response tasks, we repeated the analysis for the data converted to unbiased hit rates (Wagner, 1993). For ease of readability, pretransformed unbiased hit rate means (0–1) are reported. The same two-way repeated-measures ANOVA was conducted on observers' unbiased hit rate scores with vocal channel (2) and emotion (3) entered as within-subject factors. A main effect of vocal channel was reported, $F(1, 23) = 4.51$, $p = .045$, $\eta_p^2 = .16$, confirming that speech (M = .51, [.45, .56]) was identified more accurately than was song (M = .48, [.43, .53]). Responses also varied significantly across the emotional categories, $F(1.65, 36.88) = 44.5$, $p < .001$, $\eta_p^2 = .66$, with higher values for neutral (M = .56, [.49, .64]) and happy (M = .58, [.52, .64]) than for sad (M = .34, [.29, .39]). A two-way interaction of Channel × Emotion was also reported, $F(2, 46) = 6.15$, $p = .004$, $\eta_p^2 = .21$. Post hoc comparisons (Tukey's HSD) confirmed no differences between channels were found for happy, neutral, or sad emotions. Overall, unbiased hit rate measures replicated the results of the analyses on raw hit rates.

Discussion

Observers accurately identified the intended emotion of vocalists well above chance levels, solely on the basis of head movements presented without facial features or acoustic vocalizations. This finding confirms that head motion accompanying vocalization can convey emotional content for happy, sad, and neutral vocalizations. Observers' accuracy of emotional identification varied with the vocalized emotion: Head movements associated with happy and neutral productions were identified more accurately than were those associated with sad productions. This result suggests that some emotional head movements are more readily identifiable than others. This outcome replicates findings commonly reported in the facial and vocal expression literature, where some emotions are conveyed more accurately than others.

Vocalists' intended emotions were identified at similar levels of accuracy across speech and song, in both uncorrected (speech = 70%; song = 67%) and corrected (speech = 51%; song = 48%) measures. Although differences in accuracy for the neutral emotion were observed between speech and song, happy and sad emotions were identified equally accurately across speech and song. Overall, these results support the hypothesis that vocalists encode emotion using similar expressive head motion profiles when speaking and singing.

General Discussion

Two experiments provide converging evidence that a vocalist's emotional intent during speaking and singing is encoded in head motion. Motion capture analysis demonstrated that vocalists' head movements showed similar patterns of translational displacement and velocity and rotational movement for a given emotional intent (very happy, happy, neutral, sad, very sad) across speech and song, transcending changes in fundamental frequency that were systematic (fixed) in song but free in speech. The head movement trajectories were also comparable across four different sentences, confirming that head movements were not affected by lexical variability. Emotion-related head movements also occurred immediately prior to and after sound production, with rotational pitch (floor-or-ceiling looking) and vertical head motion beginning 3–4 s before the onset of vocalization and continuing for 2.4 s after vocalization had ended. The influence of head movements—presented without facial features or sound—on viewers' subsequent identification of emotional intent provides further evidence that head motion during speech and song functions as a visual form of emotional communication.

Vocalists' fundamental frequency (F_0) varied predictably with vocalists' emotional intent in speech, increasing for happy and very happy utterances relative to neutral, sad, and very sad vocalizations (Cowie et al., 2001; Scherer, 2003). In contrast, F_0 remained constant across emotions in song, varying only with the prescribed melody. Changes in vocalists' head movements correlated with changes in their F_0; the variance in F_0 accounted for by the six degrees of freedom of head motion was lower for song, in which F_0 is fixed by the melody, than for speech. The fact that other factors influence head movements beyond F_0 is underscored by the presence of emotionally distinct head movements in the absence of F_0 (right before or after vocalizations). The variation seen in head movements during vocalized sound may reflect orofacial movements of the jaw (Tasko & McClean, 2004; Vatikiotis-Bateson & Ostry, 1995), which are tied to the production of vocal intensity, and suprasegmental features of speech such as stress and prominence (Hadar, Steiner, Grant, & Rose, 1983; Munhall et al., 2004). Thus, vocal head movements observed here seemed to arise from at least two distinct sources—changes in acoustic parameters of the voice and changes in the expression of emotion. These results provide a novel extension to models of head motion in which linguistic and acoustic motor factors are the primary determinants of head movements (Munhall et al., 2004; Thompson & Russo, 2007; Thompson et al., 2010; Yehia et al., 2002, 1998).

Vocalists' emotions were differentiated not by a single movement trajectory but by a collection of head movement cues. Happy vocalizations exhibited larger and faster forward movement, raising of the head, and ceiling-looking rotational pitch. Conversely, sad expressions had slower forward movement, lowering of the head, and floor-looking rotational pitch. Neutral expressions exhibited attenuated movement along these dimensions. The emotional intensity of vocalizations was further differentiated by the magnitude of head movements: More-intense vocalizations generally showed larger movement patterns than did their less-intense counterparts, with very happy utterances exhibiting larger and faster forward movements and more ceiling-looking rotation relative to happy vocalizations and very sad utterances exhibiting more-forward movement and lowering of the head and greater floor-looking rotation relative to sad vocalizations. Encoding of emotion through a collection of cues, in this case the dimensions, magnitude, and direction of head movements, is also thought to be the basis of emotional communication in facial expressions, where expressions can be described through the combination of muscle *action units* (Ekman & Friesen, 1978). Similarly, emotional communication in music and speech acoustics is thought to occur through variations in acoustic cues (Juslin & Laukka, 2001, 2003; Livingstone et al., 2010; Scherer, 1995, 2003). Importantly, Experiment 2 confirmed that observers could identify vocalists' intended emotions from the presence of head movements alone. These results support the hypothesis that head movements function as part of an emotional code that is shared by encoders (vocalists) and decoders (observers;Ekman & Friesen, 1969; Elfenbein & Ambady, 2002). The use of head gestures by vocal performers may therefore function to support the emotional message conveyed by the voice, complementing the auditory signal with motion.

The present study contributes several novel findings to the understanding of emotion. First, specific patterns of head movements were found to reflect a vocalist's emotional intent for expressions of happiness and sadness. The relationship between head pose and an individual's emotional state has been posited (Atkinson et al., 2004; Darwin, 1872/1965; Matsumoto & Willingham, 2009; Wallbott, 1998), with prior work finding a link between head tilt and the nonverbal expression of self-conscious emotions such as pride and shame (Keltner, 1995; Nelson & Russell, 2014; Tracy & Robins, 2004). This is the first study to our knowledge to quantify the extent of these movement parameters, which included direction, rotation, and velocity of head motion, and that these movements distinguished happy and sad expressions of emotion, along with the intensity of these expressions. Second, we found that emotional head movements occurred before and after vocal sound production. Ceiling- and floor-looking rotational pitches and vertical movement of the head began prior to the onset of sound. When vocalization was initiated, forward movement of the head co-occurred with emotionally distinct forward-moving velocity with more-pronounced rotational pitch and vertical head movements. Although previous research has noted a relationship between rotational pitch and head motion (Atkinson et al., 2004; Boone & Cunningham, 1998), we are not aware of any reference to forward movement of the head in connection with emotion; expressive head movements accompanying vocalization may be qualitatively different from those occurring in a nonvocal context. Third, observers accurately identified a vocalist's emotional intent solely on the basis of head movements. To our knowledge, this is the first demonstration of emotion identification based solely on dynamic displays of head motion. Observers in a previous study (Ekman, 1965) identified emotions from still photographs showing either head only, body only, or head and body together. Because facial expressions were not obscured, it is unclear whether head pose contributed to observers' perception of emotion. In another study, observers identified emotions from hand-drawn images of different head orientations (Schouwstra & Hoogstraten, 1995). Because the images were not of human participants, it is unclear whether the images captured representative expressions of emotion. In more recent work, observers distinguished between displays of positive and hubristic pride from videos of dynamic facial expressions that showed the head and shoulders (Nelson & Russell, 2014). Because facial expressions were not obscured in this study, the extent to which head pose contributed to observers' emotional recognition is unknown. In contrast, the present study used video recordings of participants to capture the full range of expressive head movements, without the influence of facial features.

An ability to recognize emotional intent from head motion alone may also have applications in the area of automatic emotion recognition. With the rapid pace of development in artificial intelligence and robotics, computers are increasingly used in security situations, where video cues in the absence of audio cues is common, and as virtual or robotic agents that interact with humans, an area referred to as socially interactive robotics (Fong, Nourbakhsh, & Dautenhahn, 2003; Ishiguro, 2007). The addition of expressive vocal head movements may improve the realism of vocalizing humanoid robots and on-screen virtual avatars (Busso, Deng, Grimm, Neumann, & Narayanan, 2007), helping to close the realism gap referred to as the "uncanny valley" (Mori, MacDorman, & Kageki, 2012). That head movements encode a vocalist's emotional intentions may have applications in the area of automated affect recognition. Decoding an individual's emotional state in situations of visual noise and facial occlusion may be improved by using the information encoded in vocal head movements (Fasel & Luettin, 2003).

The present study has two notable limitations. Head movements were examined for only two emotional categories: happiness and sadness, with a neutral control condition. These emotions were selected because they exhibit low rates of confusion in perceptual studies. Happiness and sadness reflect differences in emotional valence and are thought to embody positive and negative valence, respectively (Frijda, 1986). Differences between the positive emotional conditions (very happy, happy) and negative emotional conditions (very sad, sad) may reflect aspects of emotional valence (upward for positive, downward for negative), rather than distinct categories of basic emotions. Differences in head tilt may also reflect the role of self-conscious emotions (Keltner, 1995; Nelson & Russell, 2014; Tracy & Robins, 2004). An important avenue for future work would be to investigate a broader set of discrete emotions, including anger, fear, disgust, and surprise. Because the present work reported on several novel aspects of vocal emotional head movements, including the effect of intensity on size of head movements, we theorize that a broader set of distinct emotions can be conveyed to observers by a vocalist's head movements.

A second limitation was the use of induced emotional expressions (sometimes called "simulated" or "posed" in the facial expression literature, see Bänziger et al., 2012; Gosselin, Kirouac, & Dore, 1995), instead of naturally occurring "spontaneous" displays of emotion. A criticism of instructed expressions is that they can be exaggerated, leading to inflated rates of observer agreement relative to spontaneous displays (Cohn & Schmidt, 2004; Elfenbein & Ambady, 2002; Motley & Camden, 1988; Wagner, MacDonald, & Manstead, 1986). The use of spontaneous expression paradigms may yield attenuated patterns of head movement relative to those reported in Experiment 1 or result in lower rates of observer agreement than those found in Experiment 2, or both. Induction procedures like the one used in the current study are increasingly used to obtain more-genuine expressions of emotion, while retaining the control required for a within-subject design (see also Bänziger et al., 2012; Douglas-Cowie et al., 2007; Ebner, Riediger, & Lindenberger, 2010; Kaulard et al., 2012; Livingstone et al., 2014). The comparison of induced and spontaneous displays of vocal emotional communication is an important avenue for further research. Video-based motion tracking, which does not require the use of motion capture markers, will likely be required to enable natural recordings of live music performances and real-world speech, although there remain many unsolved technical and theoretical issues when using naturalistic recordings (Douglas-Cowie, Campbell, Cowie, & Roach, 2003; Scherer, 2003).

Conclusion

In two experiments, we demonstrated that vocalists' head movements encode emotional information in speech and song and that observers could identify emotions on the basis of head movements alone. Seasoned oratory and singing performers are renowned for their dramatic and expressive use of head and facial movements. These head movements may therefore function in part to support the acoustic signal, communicating emotion through motion.

References

Ambadar, Z., Cohn, J. F., & Reed, L. I. (2009). All smiles are not created equal: Morphology and timing of smiles perceived as amused, polite, and embarrassed/nervous. *Journal of Nonverbal Behavior, 33,* 17–34. http://dx.doi.org/10.1007/s10919-008-0059-5

Atkinson, A. P., Dittrich, W. H., Gemmell, A. J., & Young, A. W. (2004). Emotion perception from dynamic and static body expressions in point-light and full-light displays. *Perception, 33,* 717–746. http://dx.doi.org/10.1068/p5096

Bänziger, T., Mortillaro, M., & Scherer, K. R. (2012). Introducing the Geneva Multimodal expression corpus for experimental research on emotion perception. *Emotion, 12,* 1161–1179. http://dx.doi.org/10.1037/a0025827

Bassili, J. N. (1978). Facial motion in the perception of faces and of emotional expression. *Journal of Experimental Psychology: Human Perception and Performance, 4,* 373–379. http://dx.doi.org/10.1037/0096-1523.4.3.373

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,* 2049–2058. http://dx.doi.org/10.1037/0022-3514.37.11.2049

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B: Methodological, 57,* 289–300.

Benjamini, Y., & Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics, 29,* 1165–1188.

Boersma, P., & Weenink, D. (2010). Praat: Doing phonetics by computer (Version 5.2.15) [Computer program]. Retrieved from http://www.praat.org/

Boone, R. T., & Cunningham, J. G. (1998). Children's decoding of emotion in expressive body movement: The development of cue attunement. *Developmental Psychology, 34,* 1007–1016. http://dx.doi.org/10.1037/0012-1649.34.5.1007

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10,* 433–436. http://dx.doi.org/10.1163/156856897X00357

Brunswik, E. (1956). *Perception and the representative design of psychological experiments*. Berkeley: University of California Press.

Bugental, D. B. (1986). Unmasking the "polite smile": Situational and personal determinants of managed affect in adult-child interaction. *Personality and Social Psychology Bulletin, 12,* 7–16. http://dx.doi.org/10.1177/0146167286121001

Busso, C., Deng, Z., Grimm, M., Neumann, U., & Narayanan, S. (2007). Rigid head motion in expressive speech animation: Analysis and synthesis. *IEEE Transactions on Audio, Speech, and Language Processing, 15,* 1075–1086.

Cacioppo, J. T., Petty, R. E., Losch, M. E., & Kim, H. S. (1986). Electromyographic activity over facial muscle regions can differentiate the valence and intensity of affective reactions. *Journal of Personality and Social Psychology, 50,* 260–268. http://dx.doi.org/10.1037/0022-3514.50.2.260

Carlo, N. S., & Guaitella, I. (2004). Facial expressions of emotion in speech and singing. *Semiotica, 2004,* 37–55. http://dx.doi.org/10.1515/semi.2004.036

Cohn, J. F., & Schmidt, K. L. (2004). The timing of facial motion in posed and spontaneous smiles. *International Journal of Wavelets, Multiresolution, and Information Processing, 2,* 121–132. http://dx.doi.org/10.1142/S021969130400041X

Coltheart, M. (1981). The MRC Psycholinguistic Database. *Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 33,* 497–505. http://dx.doi.org/10.1080/14640748108400805

Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, W., & Taylor, J. G. (2001). Emotion recognition in human-computer interaction. *IEEE Signal Processing Magazine, 18,* 32–80. http://dx.doi.org/10.1109/79.911197

Cunningham, D. W., & Wallraven, C. (2009). Dynamic information for the recognition of conversational expressions. *Journal of Vision, 9,* 7. http://dx.doi.org/10.1167/9.13.7

Dalla Bella, S., Peretz, I., Rousseau, L., & Gosselin, N. (2001). A developmental study of the affective value of tempo and mode in music. *Cognition, 80*(3), B1–B10. http://dx.doi.org/10.1016/S0010-0277(00)00136-0

Darwin, C. (1965). *The expression of emotions in man and animals*. Chicago, IL: University 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. http://dx.doi.org/10.1177/030573569302100201

Dittrich, W. H., Troscianko, T., Lea, S. E., & Morgan, D. (1996). Perception of emotion from dynamic point-light displays represented in dance. *Perception, 25,* 727–738. http://dx.doi.org/10.1068/p250727

Douglas-Cowie, E., Campbell, N., Cowie, R., & Roach, P. (2003). Emotional speech: Towards a new generation of databases. *Speech Communication, 40,* 33–60. http://dx.doi.org/10.1016/S0167-6393(02)00070-5

Douglas-Cowie, E., Cowie, R., Sneddon, I., Cox, C., Lowry, O., McRorie, M., . . . Karpouzis, K. (2007). The HUMAINE database: Addressing the collection and annotation of naturalistic and induced emotional data. In A. C. R. Paiva, R. Prada, & R. W. Picard (Eds.), *Affective Computing and Intelligent Interaction: Second international conference, ACII 2007 Lisbon, Portugal, September 12–14, 2007 proceedings* (pp. 488–500). Berlin, Germany: Springer-Verlag.

E-Prime (Version 2.0). (2012). [Computer program]. Pittsburgh, PA (http://www.pstnet.com): Psychology Software Tools.

Ebner, N. C., Riediger, M., & Lindenberger, U. (2010). FACES—A database of facial expressions in young, middle-aged, and older women and men: Development and validation. *Behavior Research Methods, 42,* 351–362. http://dx.doi.org/10.3758/BRM.42.1.351

Ekman, P. (1965). Differential communication of affect by head and body cues. *Journal of Personality and Social Psychology, 2,* 726–735. http://dx.doi.org/10.1037/h0022736

Ekman, P., & Friesen, W. V. (1969). The repertoire of nonverbal behavior: Categories, origins, usage, and coding. *Semiotica, 1,* 49–98. http://dx.doi.org/10.1515/semi.1969.1.1.49

Ekman, P., & Friesen, W. V. (1978). *Manual of the Facial Action Coding System (FACS)*. Palo Alto, CA: Consulting Psychologists Press.

Elfenbein, H. A., & Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: A meta-analysis. *Psychological Bulletin, 128,* 203–235. http://dx.doi.org/10.1037/0033-2909.128.2.203

Fasel, B., & Luettin, J. (2003). Automatic facial expression analysis: A survey. *Pattern Recognition, 36,* 259–275. http://dx.doi.org/10.1016/S0031-3203(02)00052-3

Fong, T., Nourbakhsh, I., & Dautenhahn, K. (2003). A survey of socially interactive robots. *Robotics and Autonomous Systems, 42,* 143–166. http://dx.doi.org/10.1016/S0921-8890(02)00372-X

Frijda, N. H. (1986). *The emotions*. New York, NY: Cambridge University Press.

Gabrielsson, A., & Lindström, E. (2001). The influence of musical structure on emotional expression. In P. N. Juslin & J. A. Sloboda (Eds.), *Music and emotion: Theory and research* (pp. 223–248). New York, NY: Oxford University Press.

Ghazanfar, A. A., & Takahashi, D. Y. (2014). The evolution of speech: Vision, rhythm, cooperation. *Trends in Cognitive Sciences, 18,* 543–553. http://dx.doi.org/10.1016/j.tics.2014.06.004

Gosselin, P., Kirouac, G., & Doré, F. Y. (1995). Components and recognition of facial expression in the communication of emotion by actors. *Journal of Personality and Social Psychology, 68,* 83–96. http://dx.doi.org/10.1037/0022-3514.68.1.83

Grant, K. W., & Seitz, P.-F. (2000). The use of visible speech cues for improving auditory detection of spoken sentences. *Journal of the Acoustical Society of America, 108,* 1197–1208. http://dx.doi.org/10.1121/1.1288668

Green, G. A. (1994). Unison versus individual singing and elementary students' vocal pitch accuracy. *Journal of Research in Music Education, 42,* 105–114. http://dx.doi.org/10.2307/3345495

Hadar, U., Steiner, T. J., Grant, E. C., & Rose, F. C. (1983). Head movement correlates of juncture and stress at sentence level. *Language and Speech, 26* (Pt 2), 117–129.

IBM Corp. Released. (2011). IBM SPSS Statistics for Windows, Version 20.0. Armonk, NY: IBM Corp.

Ishiguro, H. (2007). Android science. In S. Thrun, R. Brooks, & H. Durant-Whyte (Eds.), *Robotics research: Results of the 12th International Symposium ISRR* (pp. 118–127). Berlin, Germany: Springer. http://dx.doi.org/10.1007/978-3-540-48113-3_11

Jesse, A., & Massaro, D. W. (2010). Seeing a singer helps comprehension of the song's lyrics. *Psychonomic Bulletin & Review, 17,* 323–328. http://dx.doi.org/10.3758/PBR.17.3.323

Juslin, P. N. (2000). Cue utilization in communication of emotion in music performance: Relating performance to perception. *Journal of Experimental Psychology: Human Perception and Performance, 26,* 1797–1812. http://dx.doi.org/10.1037/0096-1523.26.6.1797

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. 309–340). New York, NY: Oxford University Press.

Juslin, P. N., & Laukka, P. (2001). Impact of intended emotion intensity on cue utilization and decoding accuracy in vocal expression of emotion. *Emotion, 1,* 381–412.

Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin, 129,* 770–814.

Juslin, P. N., & Västfjäll, D. (2008). Emotional responses to music: The need to consider underlying mechanisms. *Behavioral and Brain Sciences, 31,* 559–575. http://dx.doi.org/10.1017/S0140525X08005293

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. http://dx.doi.org/10.1068/p3131

Kastner, M. P., & Crowder, R. G. (1990). Perception of the major/minor distinction: IV. Emotional connotations in young children. *Music Perception, 8,* 189–201. http://dx.doi.org/10.2307/40285496

Kaulard, K., Cunningham, D. W., Bülthoff, H. H., & Wallraven, C. (2012). The MPI Facial Expression Database—A validated database of emotional and conversational facial expressions. *PLoS ONE, 7*(3), e32321. http://dx.doi.org/10.1371/journal.pone.0032321

Keltner, D. (1995). Signs of appeasement: Evidence for the distinct displays of embarrassment, amusement, and shame. *Journal of Personality* *and Social Psychology, 68,* 441–454. http://dx.doi.org/10.1037/0022-3514.68.3.441

Kohler, C. G., Turner, T., Stolar, N. M., Bilker, W. B., Brensinger, C. M., Gur, R. E., & Gur, R. C. (2004). Differences in facial expressions of four universal emotions. *Psychiatry Research, 128,* 235–244. http://dx.doi.org/10.1016/j.psychres.2004.07.003

Krumhuber, E. G., & Kappas, A. (2005). Moving smiles: The role of dynamic components for the perception of the genuineness of smiles. *Journal of Nonverbal Behavior, 29,* 3–24. http://dx.doi.org/10.1007/s10919-004-0887-x

Krumhuber, E. G., Kappas, A., & Manstead, A. S. R. (2013). Effects of dynamic aspects of facial expressions: A review. *Emotion Review, 5,* 41–46. http://dx.doi.org/10.1177/1754073912451349

Ladefoged, P., & Johnson, K. (2010). *A course in phonetics*. Boston, MA: Wadsworth.

Livingstone, S. R., Choi, D. H., & Russo, F. A. (2014). The influence of vocal training and acting experience on measures of voice quality and emotional genuineness. *Frontiers in Psychology, 5,* 156. http://dx.doi.org/10.3389/fpsyg.2014.00156

Livingstone, S. R., Muhlberger, R., Brown, A. R., & Thompson, W. F. (2010). Changing musical emotion: A computational rule system for modifying score and performance. *Computer Music Journal, 34,* 41–64. http://dx.doi.org/10.1162/comj.2010.34.1.41

Livingstone, S. R., Palmer, C., & Schubert, E. (2012). Emotional response to musical repetition. *Emotion, 12,* 552–567. http://dx.doi.org/10.1037/a0023747

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. http://dx.doi.org/10.1525/mp.2009.26.5.475

Livingstone, S. R., Thompson, W. F., Wanderley, M. M., & Palmer, C. (2015). Common cues to emotion in the dynamic facial expressions of speech and song. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 68,* 952–970.

Matlab (2013) [Computer software]. Natick, MA: MathWorks.

Matsumoto, D., & Willingham, B. (2009). Spontaneous facial expressions of emotion of congenitally and noncongenitally blind individuals. *Journal of Personality and Social Psychology, 96,* 1–10. http://dx.doi.org/10.1037/a0014037

McGurk, H., & MacDonald, J. (1976, December 23). Hearing lips and seeing voices. *Nature, 264,* 746–748. http://dx.doi.org/10.1038/264746a0

Mori, M., MacDorman, K. F., & Kageki, N. (2012). The uncanny valley [from the field]. *IEEE Robotics & Automation Magazine, 19,* 98–100. http://dx.doi.org/10.1109/MRA.2012.2192811

Motley, M. T., & Camden, C. T. (1988). Facial expression of emotion: A comparison of posed expressions versus spontaneous expressions in an interpersonal communication setting. *Western Journal of Speech Communication, 52,* 1–22. http://dx.doi.org/10.1080/10570318809389622

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. http://dx.doi.org/10.1111/j.0963-7214.2004.01502010.x

Nelson, N. L., & Russell, J. A. (2014). Dynamic facial expressions allow differentiation of displays intended to convey positive and hubristic pride. *Emotion, 14,* 857–864. http://dx.doi.org/10.1037/a0036789

O'Toole, A. J., Roark, D. A., & Abdi, H. (2002). Recognizing moving faces: A psychological and neural synthesis. *Trends in Cognitive Sciences, 6,* 261–266. http://dx.doi.org/10.1016/S1364-6613(02)01908-3

Otta, E., Lira, B. B. P., Delevati, N. M., Cesar, O. P., & Pires, C. S. G. (1994). The effect of smiling and of head tilting on person perception. *Journal of Psychology, 128,* 323–331. http://dx.doi.org/10.1080/00223980.1994.9712736

Plant, R. R., Hammond, N., & Whitehouse, T. (2002). Toward an Experimental Timing Standards Lab: Benchmarking precision in the real world. *Behavior Research Methods, Instruments, & Computers, 34,* 218–226. http://dx.doi.org/10.3758/BF03195446

Quinto, L. R., Thompson, W. F., Kroos, C., & Palmer, C. (2014). Singing emotionally: A study of pre-production, production, and post-production facial expressions. *Frontiers in Psychology, 5,* 262. http://dx.doi.org/10.3389/fpsyg.2014.00262

Ramsay, J. O., & Silverman, B. W. (Eds.). (2002). *Applied functional data analysis: Methods and case studies* (Vol. 77). New York, NY: Springer. http://dx.doi.org/10.1007/b98886

Ramsay, J. O., & Silverman, B. W. (2005). *Functional data analysis*. New York, NY: Springer. http://dx.doi.org/10.1002/0470013192.bsa239

Ramsay, J. O., & Silverman, B. W. (2010). *Functional data analysis package*. Retrieved from http://www.psych.mcgill.ca/misc/fda/software.html

Scherer, K. R. (1995). Expression of emotion in voice and music. *Journal of Voice, 9,* 235–248. http://dx.doi.org/10.1016/S0892-1997(05)80231-0

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

Scherer, K. R., Banse, R., Wallbott, H. G., & Goldbeck, T. (1991). Vocal cues in emotion encoding and decoding. *Motivation and Emotion, 15,* 123–148. http://dx.doi.org/10.1007/BF00995674

Schouwstra, S. J., & Hoogstraten, J. (1995). Head position and spinal position as determinants of perceived emotional state. *Perceptual and Motor Skills, 81,* 673–674. http://dx.doi.org/10.2466/pms.1995.81.2.673

Tasko, S. M., & McClean, M. D. (2004). Variations in articulatory movement with changes in speech task. *Journal of Speech, Language, and Hearing Research, 47,* 85–100. http://dx.doi.org/10.1044/1092-4388(2004/008)

Thompson, W. F., & Russo, F. A. (2007). Facing the music. *Psychological Science, 18,* 756–757. http://dx.doi.org/10.1111/j.1467-9280.2007.01973.x

Thompson, W. F., Russo, F. A., & Livingstone, S. R. (2010). Facial expressions of singers influence perceived pitch relations. *Psychonomic Bulletin & Review, 17,* 317–322. http://dx.doi.org/10.3758/PBR.17.3.317

Tracy, J. L., & Matsumoto, D. (2008). The spontaneous expression of pride and shame: Evidence for biologically innate nonverbal displays. *PNAS Proceedings of the National Academy of Sciences of the United States of America, 105,* 11655–11660. http://dx.doi.org/10.1073/pnas.0802686105

Tracy, J. L., & Robins, R. W. (2004). Show your pride: Evidence for a discrete emotion expression. *Psychological Science, 15,* 194–197. http://dx.doi.org/10.1111/j.0956-7976.2004.01503008.x

Van den Stock, J., Righart, R., & de Gelder, B. (2007). Body expressions influence recognition of emotions in the face and voice. *Emotion, 7,* 487–494. http://dx.doi.org/10.1037/1528-3542.7.3.487

Vatikiotis-Bateson, E., & Ostry, D. J. (1995). An analysis of the dimensionality of jaw motion in speech. *Journal of Phonetics, 23,* 101–117. http://dx.doi.org/10.1016/S0095-4470(95)80035-2

Vines, B. W., Krumhansl, C. L., Wanderley, M. M., Dalca, I. M., & Levitin, D. J. (2011). Music to my eyes: Cross-modal interactions in the perception of emotions in musical performance. *Cognition, 118,* 157–170. http://dx.doi.org/10.1016/j.cognition.2010.11.010

Wagner, H. L. (1993). On measuring performance in category judgment studies of nonverbal behavior. *Journal of Nonverbal Behavior, 17,* 3–28. http://dx.doi.org/10.1007/BF00987006

Wagner, H. L., MacDonald, C. J., & Manstead, A. (1986). Communication of individual emotions by spontaneous facial expressions. *Journal of Personality and Social Psychology, 50,* 737–743. http://dx.doi.org/10.1037/0022-3514.50.4.737

Wallbott, H. G. (1998). Bodily expression of emotion. *European Journal of Social Psychology, 28,* 879–896. http://dx.doi.org/10.1002/(SICI)1099-0992(1998110)28:6<879::AID-EJSP901>3.0.CO;2-W

Wallbott, H. G., & Scherer, K. R. (1986). Cues and channels in emotion recognition. *Journal of Personality and Social Psychology, 51,* 690–699. http://dx.doi.org/10.1037/0022-3514.51.4.690

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. http://dx.doi.org/10.1080/09298210500124208

Wiener, M., Devoe, S., Rubinow, S., & Geller, J. (1972). Nonverbal behavior and nonverbal communication. *Psychological Review, 79,* 185–214. http://dx.doi.org/10.1037/h0032710

Yehia, H. C., Kuratate, T., & Vatikiotis-Bateson, E. (2002). Linking facial animation, head motion and speech acoustics. *Journal of Phonetics, 30,* 555–568. http://dx.doi.org/10.1006/jpho.2002.0165

Yehia, H. C., Rubin, P., & Vatikiotis-Bateson, E. (1998). Quantitative association of vocal-tract and facial behavior. *Speech Communication, 26,* 23–43. http://dx.doi.org/10.1016/S0167-6393(98)00048-X

Received March 25, 2015
Revision received June 16, 2015
Accepted June 24, 2015 ■