The Acoustic Correlates of Valence Depend on Emotion Family

Michel Belyk and Steven Brown, *Hamilton, Ontario, Canada*

Summary: The voice expresses a wide range of emotions through modulations of acoustic parameters such as frequency and amplitude. Although the acoustics of individual emotions are well understood, attempts to describe the acoustic correlates of broad emotional categories such as valence have yielded mixed results. In the present study, we analyzed the acoustics of emotional valence for different families of emotion. We divided emotional vocalizations into "motivational," "moral," and "aesthetic" families as defined by the OCC (Ortony, Clore, and Collins) model of emotion. Subjects viewed emotional scenarios and were cued to vocalize congruent exclamations in response to them, for example, "Yay!" and "Damn!". Positive valence was weakly associated with high-pitched and loud vocalizations. However, valence interacted with emotion family for both pitch and amplitude. A general acoustic code for valence does not hold across families of emotion, whereas family-specific codes provide a more accurate description of vocal emotions. These findings are consolidated into a set of "rules of expression" relating vocal dimensions to emotion dimensions.

Key Words: Prosody–Speech–Acoustics–Vocalization–Valence–Pitch–Cognitive appraisal–Emotion.

INTRODUCTION

The voice expresses a wide range of emotions through modulations of acoustic parameters such as frequency and amplitude. Although there is an extensive literature devoted to the acoustic analysis of vocal expression,[1,2] this literature has tended to be descriptive rather than analytical. In other words, it has tended to look at the acoustic patterns of particular emotions without attempting to create an overarching framework for the analysis of all emotions. The challenge of creating an acoustic code of emotional expression depends on relating the dimensions of emotion to the dimensions of the voice. In the domain of animal vocalization, a synthetic "motivation-structural code" relates, for example, the emotional states of aggressiveness and fear to the vocal dimensions of pitch and timbre.[3,4] A similar set of expression rules is needed for human emotional expression.

Numerous studies of affective speech prosody have shown that frequency (the physical correlate of pitch) and amplitude (the physical correlate of loudness) are the two acoustic features that vary most strongly across emotions[5] and that either cue in isolation is sufficient for above-chance discrimination of prosody.[6] Hence, these two vocal dimensions might be the best place to start in thinking about expression rules in humans. Early studies of vocal emotion focused primarily on measures related to pitch.[7] More recently, several large-scale corpora have provided descriptive data for a variety of emotional expressions within the context of language[5,8] as well as for exclamations outside the context of language.[9] These studies have demonstrated the relatively strong affective signal found in pitch and amplitude. These acoustic features are used in a similar fashion by speakers of different languages[10] and form part of the basis for the discrimination of relatively subtle emotional distinctions, such as that between joyous laughter and taunting laughter.[11] In addition, sensitivity to pitch and amplitude cues in speech appears to develop early in infancy. Mothers use infant-directed speech, as characterized by a raised vocal pitch,[12] in their communication with babies, and infants attend preferentially to this form of speech over adult-directed speech.[13] Children use pitch and amplitude (among other cues) differentially so as to vary the form and intensity of their tantrums.[14] Furthermore, the use of vocal pitch as an affective cue is highly preserved across species.[15,16] However, the salience of pitch and amplitude by no means negates the importance of other acoustic cues. Voice quality, such as breathy or creaky voice, also serves as an informative affective cue.[17,18]

Although pitch and amplitude are salient acoustic dimensions to examine for vocal expression, which dimensions of emotion should be examined? This question hinges on which emotion theory is adopted, as the theories classify emotions in disparate manners. "Basic emotions theory" (BET) predicts that there should be a distinct acoustic pattern for each emotion, analogous to the distinct patterns of facial expression that form the empirical foundations of this theory.[19] Although this prediction has been supported by both acoustic[5] and perceptual[20,21] studies, BET makes no predictions regarding acoustic relationships among the emotions because it sees all emotions as singular types. Hence, BET amounts to a null hypothesis of no relationship among emotions. In contrast to this, several contemporary emotion theories function as umbrella models that attempt to encompass all emotions in a dimensional manner. "Core affect theory"[22,23] posits that any emotion can be situated somewhere along the two orthogonal dimensions of valence and arousal. As discussed below, there is strong evidence for the vocal encoding of arousal but only weak and mixed evidence for the encoding of valence.

Cognitive appraisal theories of emotion sit at an intermediate position between the extremes of emotions as singular types (as in BET) or as points along a set of independent continua (as in core affect theory). These approaches generally include

---

Accepted for publication December 11, 2013.

This work was funded by a grant from the Natural Sciences and Engineering Research Council (NSERC) of Canada to S.B.

From the Department of Psychology, Neuroscience and Behaviour, McMaster University, Hamilton, Ontario, Canada.

Address correspondence and reprint requests to Michel Belyk, Department of Psychology, Neuroscience & Behaviour, McMaster University, 1280 Main St. West, Hamilton, Ontario L8S 4K1, Canada. E-mail: belykm@mcmaster.ca

Journal of Voice, Vol. 28, No. 4, pp. 523.e9-523.e18

0892-1997/$36.00

http://dx.doi.org/10.1016/j.jvoice.2013.12.007

well-established concepts—such as emotional valence—as well as classes of emotions based on the context in which the emotion is elicited.

A lesser-known appraisal theory of emotion—one that informs the present study—is the "cognitive structure model" put forth by Ortony, Clore, and Collins[24]; hereafter, the OCC model. Although this model contains the valence and arousal dimensions of core affect theory, it adds a third factor of "emotion family." In particular, the model posits the existence of three emotion families based on the cognitive appraisals that lead to their elicitation. (1) "Motivational" emotions (or what OCC refers to as "outcome" emotions) result from the appraisal of the value of an outcome relative to a set of personal goals (eg, joy vs distress). (2) "Moral" emotions ("agency" emotions in OCC) result from the appraisal of the behavior of people relative to a set of social standards (eg, gratitude vs anger). (3) "Aesthetic" emotions ("object" emotions in OCC) result from the appraisal of an object's appeal relative to an internalized set of desirable properties (eg, pleasure vs disgust). Each appraisal type can produce an emotion of either positive or negative valence, hence, leading to a structure in which emotions are organized into opponent valence pairs (eg, happy vs sad).

The OCC model has strong similarities to a well-known appraisal model of emotion, namely Scherer's (1986) "component process model." In fact, the latter model's five components more or less encompass the three emotion families of the OCC model but add two more distinctions: (1) "novelty detection," which is related to affective intensity; (2) "intrinsic pleasantness," which is related to the appraisal of aesthetic value; (3) "goal/need conduciveness," which is related to the appraisal of outcomes; (4) "coping potential," which has no analogue in the OCC model; and (5) "norm/self-compatibility," which is related to the appraisal of agency. The component process model, therefore, provides striking similarities to the OCC framework used in the present study.

The overarching goal of this work is less to test out competing models of emotion as to create a framework for systematizing the relationship between emotion dimensions and vocal dimensions in human expression, just as has been done for nonhuman animals.[4] In other words, there is a need to elucidate "rules of expression" that influence the production and perception of different emotion types. Rules of expression refer here to a set of expressive conventions for mapping vocal acoustics onto emotional states. These conventions may be formed by evolutionary selection pressures,[25] culture,[26] or mechanical constraints of the larynx and vocal tract.[27] Hence, expression rules may describe either external "push" effects or internal "pull" effects.[28]

The search for such rules has been successfully extended to include musical expression as well.[1,29] In both speech and music, there is a "high-loud" expression rule whereby an increase in the frequency of the voice is accompanied by an increase in amplitude.[29,30] In other words, high-frequency vocalizations tend to be loud, whereas low-frequency vocalizations tend to be soft. However, pitch and amplitude can be shown to part company as well. Fonagy[31] pointed out that although the expression of joy involves simultaneous increases in pitch and amplitude, the expression of coquettishness involves a coupling of high pitch with low amplitude.

A high-loud expression rule relates two vocal dimensions to one another, but what about voice/emotion relationships? To a first approximation, a formal analysis of rules of expression should relate the vocal dimensions of pitch and amplitude to the emotional dimensions of valence and arousal. Arousal relationships have been demonstrated quite consistently in the literature. Strong evidence supports a direct relationship between emotional arousal and both pitch height and amplitude.[32–34] The larger challenge is to explain the influence of *valence* on vocal acoustics, as this relationship has received less-conclusive support.

As an intuitive example of valence coding, consider the everyday situation of a group of sports fans attending a match at the local stadium. They will generally scream out a high-pitched "Yay!" every time the home team scores a goal but groan a lower-pitched "Aw!" when the competing team scores. There is an intuitive sense that positive emotions are vocalized with high pitch and negative emotions with low pitch, conforming to a "positive-high" rule. However, unlike arousal, the relationship between pitch height and valence has found only weak support in the experimental literature. For example, Ilie and Thompson[32] obtained affective ratings of music and speech excerpts. Although they observed a consistent positive association between arousal and sound intensity, the associations between pitch height and emotional valence were in opposite directions for music and speech. For the speech excerpts, a high pitch was associated with positive valence, whereas for the music excerpts, a high pitch was associated with negative valence. Goudbeek and Scherer[33] analyzed a corpus of acted emotional vocalizations that included emotions varying systematically in arousal, valence, and potency/control. Both pitch and sound intensity were associated with arousal but neither was associated with valence independent of arousal. Laukka et al[34] similarly obtained affective ratings from a corpus of acted emotional vocalizations. Again, the authors observed a strong positive association between arousal and both pitch and sound intensity. In contradiction to the speech corpus analysis of Ilie and Thompson,[32] valence was associated negatively with both pitch and amplitude.

Similar uncertainty about valence coding comes from Banse and Scherer,[5] who provided detailed acoustic descriptions for vocal expressions of 14 emotions. Of the four positive emotions that they included in their analysis (elation, happiness, pride, and interest), only elation was associated with a high-pitched and high-amplitude vocalization. However, their data did provide support for the association of these acoustic variables with emotional intensity. Vocal expressions of elation were both higher and louder than those of happiness. The same was seen for panic in relation to anxiety and for despair in relation to sadness. Similarly, studies explicitly investigating affective intensity have demonstrated its association with increases in both pitch and amplitude.[35,36]

One possible explanation for the lack of support for an effect of valence on vocal pitch—one that is explored in the present study—is that the direction of the acoustic effect varies across

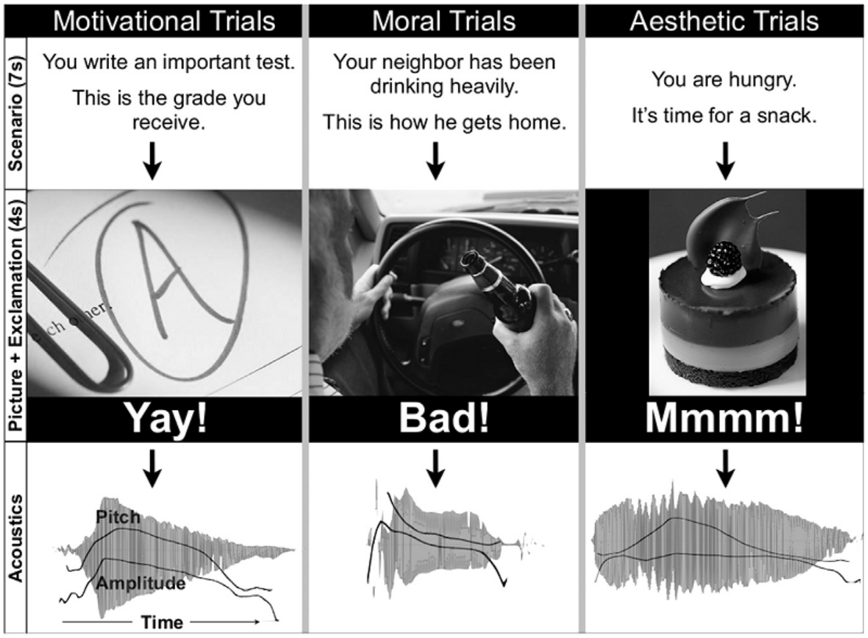

FIGURE 1. Sample trials for each of the three families of emotions. A neutral two-sentence text-scenario is presented for 7 seconds to provide a context for the emotion (top row). A valenced picture is then presented for 4 seconds to conclude the scenario (middle row). It is accompanied by a congruent exclamation-word, which the subject vocalizes. The vocalization is recorded and then analyzed for its acoustic properties (bottom row). A representative waveform as well as representative pitch and amplitude contours are shown for each family.

families of emotions, for example in the manner posited by the OCC model. In other words, although a "positive-high" rule may govern vocal expression within one particular family of emotions, other expression rules may be needed to describe valence patterns for other families. For example, the prosody of aesthetic emotions can be quite different from that of cheering fans at a sporting event. The "Mmmm!" of pleasure can occur with far lower pitch than the "Ewww!" of disgust, demonstrating the *inverse* of a "positive-high" relationship. Hence, an emotion theory like OCC that contains emotion families and valence pairings might provide novel insights into the inconclusive valence effect seen in the acoustics literature.

The principal objective of the present study was to disambiguate the acoustic correlates of emotional valence. In the experimental paradigm, subjects viewed emotional scenarios and were then recorded as they produced cued, but unrehearsed, emotional exclamations congruent with the scenarios. Nonemotional recordings of the same exclamation words served as a control for the phonemic content of the words. Two predictions were made based on the published literature, (1) a main effect of valence on vocal acoustics would either be absent or only weakly present and (2) valence would demonstrate an interaction with emotion family such that valence effects on vocal acoustics would vary as a function of emotion family.

METHODS

Subjects

Thirty undergraduates (22 females), all of them native speakers of English, were recruited from the psychology department at McMaster University. They participated in the experiment in exchange for either course credit or $10 remuneration.

Stimuli

Each experimental stimulus consisted of an emotionally neutral scenario written out in two sentences followed by an emotionally valenced picture providing a conclusion to the scenario. The picture was accompanied by a congruent exclamation-word. For example, the scenario "You write an important test. This is the grade you receive." could be followed by a picture of an examination with an A written on it accompanied by the exclamation-word "Yay!" or a picture of an examination with an F written on it accompanied by the exclamation-word "Damn!". Subjects were instructed to imagine that the scenario was happening to them and vocalize the exclamation-word in an authentic manner. Figure 1 presents some sample stimuli, and Table 1 provides definitions of the emotions included in the study. Table 2 provides the family and valence designations for each emotion as well as the exclamation words used for their expression.

The text-scenarios were designed to specify the appraisal types associated with each of the three emotion families described in the OCC model, whereas the pictures were designed to convey the valence of the outcome. Motivational scenarios contained themes related to gain or loss (eg, getting a raise or breaking a laptop), moral scenarios contained themes related to prosocial or antisocial behavior (eg, helping a homeless person or drinking alcohol while driving), and aesthetic scenarios contained themes related to encounters with desirable or undesirable food items or locations (eg, chocolate cake or a filthy toilet). Wherever possible, a given text-scenario was

TABLE 1.
Definitions of Emotion Labels in the Present Study

Emotion	Definition
Joy	Being pleased about a desirable outcome
Distress	Being displeased about an undesirable outcome
Gloating	Being pleased about an outcome presumed to be undesirable for another person
Resentment	Being displeased about an outcome presumed to be desirable for another person
Appreciation	Approving of another person's praiseworthy action
Reproach	Disapproving another person's blameworthy action
Gratitude	Approving of another person's praiseworthy action that has a desirable consequence for oneself
Anger	Disapproving of another person's blameworthy action that has an undesirable consequence for oneself
Pleasure	Liking an appealing object
Disgust	Disliking an unappealing object
Awe	Being pleased with the vastness of an object
Terror	Being displeased with the vastness of an object

Notes: The table lists the 12 emotion labels used in the present study and provides definitions based on the appraisals antecedent to each, as specified in the OCC model.

paired with both positive and negative pictures so as to create matched valence-pairs. Piloting suggested that some scenarios (as in the middle panel of Figure 1) could not effectively be paired with a picture to represent the emotion of opposite valence.

Procedure

Subjects first completed a demographic survey detailing their language background and musical training. They were then seated in a sound booth in front of a desktop computer. Recordings were made using an Apex 275 (Apex Electronics, Canada) condenser headset microphone with gain settings adjusted for each subject.

TABLE 2.
A Listing of the Emotions and Exclamations Analyzed in This Study

Family	Valence	Emotion	Exclamation
Motivational	Positive	Joy	Yay! Cool!
	Negative	*Distress*	*Damn!* *No!*
	Positive	Gloating	Ha!
	Negative	*Resentment*	*Damn!*
Moral	Positive	Appreciation	Good!
	Negative	*Reproach*	*Bad!*
	Positive	Gratitude	Thanks!
	Negative	*Anger*	*Jerk!*
Aesthetic	Positive	Pleasure	Oooh! Mmmm!
	Negative	*Disgust*	*Ewww!* *Yuck!*
	Positive	Awe	Wow!
	Negative	*Terror*	*Whoa!*

Notes: The table lists the 12 emotions examined in the present study, as well as their emotion family and valence, as specified in the cognitive structure model of emotion. The right column lists the exclamation words that were used to study the expression of each emotion. Wherever possible, more than one exclamation-word was used per emotion (15 total). Note that the emotions are organized in the table as valence pairs, in which negative emotions are shown with italics font.

To assess the basic characteristics of each subject's voice, subjects were recorded while they performed a number of vocal warm-up tasks: reading the standard "rainbow passage;" coughing; clearing the throat; performing vocal sweeps to the top and bottom of their vocal range; and singing "Happy Birthday." Each subject's habitual vocal frequency was estimated from the reading passage.

After receiving instructions, subjects performed two practice trials followed by 72 randomly ordered experimental trials (6 trials × 12 emotions). Figure 1 demonstrates the trial structure and provides example stimuli. On each trial, subjects silently read a neutral text-scenario for 7 seconds and then viewed a valenced picture for 4 seconds. The picture was accompanied by an exclamation-word, which the subjects were instructed to produce aloud. After vocalizing the exclamation, subjects were immediately asked to rate whether the scenario was pleasant or unpleasant using the "+" and "−" keys, respectively, on the keyboard as well as to rate how intense the scenario was by using the number-keys 1 (low intensity) through 9 (high intensity). Trials were separated by a 2-second rest period. All trials were carried out during a single run lasting approximately 25 minutes.

Next, during the control condition, subjects were asked to vocalize the same exclamation words but in a nonemotional speaking voice. No text-scenarios or pictures appeared in the control condition. Sixty trials (four replications of each of the 15 exclamation words) occurred in random order.

Analysis

Acoustic measures were extracted from recordings with an in-house *Praat*[37] (www.fon.hum.uva.nl/praat/) script. The voiced portion of each recorded vocalization was selected manually for automatic extraction of acoustic measures. The two measurements were (1) mean frequency relative to a subject's habitual frequency, as measured in cents (where 100 cents = 1 semitone) and (2) mean amplitude, as measured in decibels (dB). Acoustic features were measured across the entire voiced segment for each vocalization. As mentioned above, the mean frequency during the reading of the rainbow passage was used as the subject's habitual frequency and was set as the reference value in converting Hz values to cents for the pitch measurements.

The present study diverges from normative practice by reporting pitch data in semitones rather than fundamental frequency (F_0) in Hz. The difference limen for F_0 discrimination increases approximately logarithmically rather than linearly.[38] The semitone scale is a relative log-transform of F_0 that provides a better approximation of pitch perception. Furthermore, the modulation of F_0 by female speakers occurs on a larger scale than modulation of F_0 by male speakers. For example, a 100 Hz increase in F_0 above habitual pitch is of perceptually greater magnitude for a male (approximately 12 semitones) than a female (approximately 7 semitones). Conversely, producing an equivalent change in pitch requires a much larger increase in F_0 by the female. Therefore, using F_0 can result in female speakers having a disproportionate influence on the group data when collapsing across sexes.

The manual selection of vocalizations served as a quality control procedure. Algorithms that track F_0 are prone to errors, which can be corrected manually. For example, the voice quality phenomenon known as "creaky voice" tends to bias these algorithms toward implausible F_0 values. Wherever creaky voice was detected, the analyst (M.B.) attempted to correct for it by adjusting parameters to the frequency-tracking algorithm until a plausible F_0 value was obtained. If a plausible estimate of F_0 could not be obtained, the implausible segment of the sound was omitted so as to exclude it from further analysis while still preserving the integrity of the sound intensity measurements. If the implausible segment could not be successfully removed without distorting the remaining plausible segment of the vocalization, the trial was discarded.

To remove the effect of segmental structure on the vocal acoustics of the exclamations,[39] the pitch values of the control trials were subtracted from the pitch values of the corresponding experimental trials. These "difference scores" were calculated by subtracting the within-subject average profile of each control word from each experimental trial in which that exclamation-word was uttered. For example, for every experimental trial in which a subject exclaimed "Yay!", the subject's average pitch for the neutrally produced "Yay" was subtracted. On the assumption that differences among control words are due to phonetic effects and that the influence of segmental structure is equivalent in affective and nonaffective contexts, then a difference-score approach should remove the confounding effects of segmental structure on vocal acoustics. During debriefing, nine of 30 subjects reported difficulty in producing at least some of the exclamation words in a neutral fashion. To the extent that subjects may have uttered the exclamation words in an emotional rather than neutral fashion during control recordings, the difference-score method would have had the effect of underestimating some of the prosodic features of the exclamations. This subtraction method was applied to pitch only, as an analysis of subjects' control recordings revealed that pitch, but not amplitude, was affected by segmental structure.

Linear mixed models were used to test six *a priori* directional hypotheses based on pilot data, using maximum likelihood estimation in R[40] with the lme4 package.[41] For each family of emotions, tests were conducted to compare positive versus negative emotions for both pitch and amplitude. Six additional *post hoc* Bonferroni-corrected tests were conducted to examine whether the emotion-pairs that make up the aesthetic family followed qualitatively similar trends. In all cases, affective intensity ratings were included as a continuous covariate.

Examination of the data set showed that there were no statistical outliers among the subjects and that the residuals from the regression analyses were normally distributed. The group results are thus representative of the trends seen in the individual subjects.

RESULTS

Affective intensity

Subjects' ratings of affective intensity were examined to determine whether affective intensity was equivalent across emotion families and valences. Subjects' ratings of affective intensity differed across valences, $F(1, 354) = 5.35$, $P < 0.05$ with effect size Cohen $f = 0.11$. Negative emotions were rated as more intense, on average, than positive emotions. Affective intensity ratings also differed across emotion families, $F(2, 354) = 29.67$, $P < 0.05$ with Cohen $f = 0.15$. The valence-by-family interaction was not significant, $F(2, 354) = 0.95$, $P = 0.38$. Aesthetic emotions were rated as more intense than both moral emotions, $t(238) = 5.67$, $P < 0.05$ with Cohen $d = 0.19$ and motivational emotions, $t(238) = 6.59$, $P < 0.05$ with Cohen $d = 0.23$. Motivational and moral emotions were not rated differently from one another, $t(238) = 1.09$, $P = 0.28$. Although these effects were small, the results indicated that affective intensity was not fully equated across the conditions. For this reason, affective-intensity ratings were used as covariates with type III sums-of-squares in all analyses.

Phonetic control

To test whether the segmental content of the exclamation words had a measurable effect on vocal acoustics, separate univariate analyses of variance were performed on the pitch and amplitude measurements of the neutrally produced exclamation words. This analysis turned out to be significant for pitch, $F(14, 435) = 2.01$, $P < 0.05$, but not for amplitude, $F(14, 435) = 0.92$, $P = 0.54$. For example, neutrally produced "Bad" was produced with a lower pitch than neutrally produced "Yuck". For this reason, the difference-score approach described in the Methods section was applied to the pitch data (only) to eliminate the effects of segmental structure on the acoustic results.

Acoustic profiles of the emotion families

In contrast to much of the literature dealing with the coding of emotional valence in acoustic prosody, there was a main effect of valence on both pitch, $F(1, 353) = 13.5$, $P < 0.05$, and

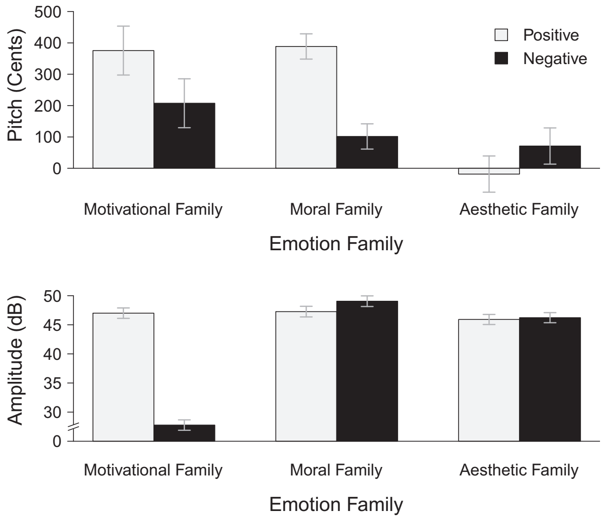

FIGURE 2. Acoustic properties of the produced exclamations. The top panel provides the pitch results (in cents relative to subjects' habitual pitch) and the bottom panel provides the amplitude results (in decibels) for the three families of emotions in the OCC model. Unfilled bars are positive-valence emotions and filled bars are negative-valence emotions. Error bars are 95% confidence intervals. The y-axis for amplitude is minimized to visualize the error bars and the differences between means.

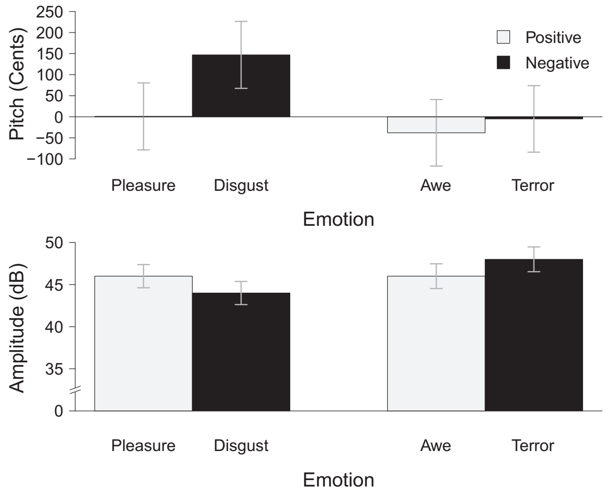

FIGURE 3. Acoustic properties of the aesthetic exclamations alone. The top panel provides the pitch results (in cents relative to subjects' habitual pitch) and the bottom panel provides the amplitude results (in decibels) for the four emotions that made up the aesthetic family of emotions. Bars are grouped such that the unambiguously aesthetic emotions (pleasure and disgust) are on the left of each plot, whereas the so-called numinous emotions are on the right. Error bars are 95% confidence intervals. The y-axis for amplitude is minimized to visualize the error bars and the differences between means.

amplitude, $F(1, 353) = 12.2$, $P < 0.05$. Positive emotions were vocalized with both higher pitch and higher amplitude than negative emotions when collapsing across the entire stimulus set. A main effect of emotion family was also observed for both pitch $F(2, 353) = 24.2$, $P < 0.05$ and amplitude $F(2, 353) = 10.6$, $P < 0.05$. There was a significant interaction between valence and emotion family for both pitch $F(2, 353) = 14.2$, $P < 0.05$ and amplitude $F(2, 353) = 12.4$, $P < 0.05$. It is, therefore, inappropriate to interpret a linear relationship between either acoustic dimension and emotional valence. The influence of emotional valence on vocal pitch and loudness may differ between emotion categories.

Given this significant interaction, the next step was to examine the acoustic profiles within the three emotion families specified in the OCC model. Figure 2 presents the pitch and amplitude profiles for the three emotion families individually. For motivational emotions, positive emotions were both higher in pitch, $t(118) = 4.24$, $P < 0.05$ with Cohen $d = 0.54$ and greater in amplitude, $t(118) = 5.42$, $P < 0.05$ with Cohen $d = 0.78$ than negative emotions. For moral emotions, however, a different profile emerged. Although positive emotions were significantly higher in pitch than negative emotions, $t(118) = 6.87$, $P < 0.05$ with Cohen $d = 0.99$, they were also significantly lower in amplitude than negative emotions, $t(118) = -1.88$, $P < 0.05$ with Cohen $d = 0.19$. Finally, aesthetic emotions presented an additional reversal from the pattern of the motivational emotions. Positive emotions were significantly lower in pitch than negative emotions, $t(118) = -3.317$, $P < 0.05$ with Cohen $d = 0.38$ and significantly lower in amplitude, $t(118) = -5.79$, $P < 0.05$ with Cohen $d = 0.32$ than negative emotions.

To further explore the reversal of the "high-loud" pattern observed for the aesthetic emotions, these emotions were divided into two subtypes: pleasure/disgust and awe/terror. Six *post hoc* comparisons were conducted to test for differences in the means of the two emotion-pairs for both pitch and amplitude, with an alpha threshold of .008 after Bonferroni correction for multiple comparisons. Figure 3 presents the acoustic profiles for each aesthetic emotion. Pleasure and disgust displayed the same overall pattern as the aesthetic analysis for pitch, $t(58) = -3.59$, $P < 0.008$ with Cohen $d = 0.64$, and amplitude, $t(58) = -4.49$, $P = 0.83$ with Cohen $d = 0.65$. However, awe and terror displayed a qualitatively different pattern. There was no significant difference between them for pitch, $t(58) = -0.945$ $P = 0.008$. However, awe was significantly lower in amplitude than terror, $t(58) = -2.9$, $P < 0.008$ with Cohen $d = 0.44$. This result may suggest that awe and terror—referred to as "numinous" emotions in the Discussion section—do not form a cohesive group with pleasure and disgust but may instead belong to a separate subfamily of emotions. It also shows that the overall profile for the aesthetic emotions was driven by pleasure/disgust.

Although the disgust-related stimuli and vocal expressions in the present study were uniformly related to disease,[42,43] the pleasure-related stimuli and expressions included both gustatory ("Mmmm!") and nongustatory ("Oooh!") forms. Therefore, the profiles of gustatory and nongustatory vocalizations were compared. Vocal expressions of gustatory and nongustatory pleasure were not significantly different in pitch, $t(58) = 1.354$, $P = 0.16$. However, vocal expressions of nongustatory pleasure ("Oooh!") were significantly higher in amplitude than those of gustatory pleasure ("Mmmm!"), $t(58) = 6.06$, $P < 0.008$ Cohen $d = 1.02$.

DISCUSSION

The present study sought to relate the vocal dimensions of pitch and amplitude to the emotional dimension of valence. We found a weak but significant effect of valence on pitch and amplitude. Previous studies that have examined the acoustic coding of emotional valence have provided inconsistent results regarding the intuitive notion that positive emotions are expressed with high-pitched vocalizations and negative emotions with low-pitched vocalizations.[32–34,44] The results of the present study suggest that this inconsistency may be due to idiosyncrasies in the choice of emotions included in each study.

Grouping vocal expressions into broad emotion families revealed that the acoustic valence-code varies across families. In particular, opposing effects were seen among the emotion families, and this could potentially explain the inconsistent findings of previous studies. Only the motivational family of emotions (eg, happiness vs sadness) showed the expected coupling between pitch, amplitude, and valence.[30] The moral family demonstrated an inverted amplitude trend such that positive moral emotions (eg, gratitude) were associated with low amplitude, whereas negative moral emotions (eg, anger) were associated with high-amplitude vocalizations. This profile is consistent with an ethological perspective in which aggressive vocalizations with low pitch and high amplitude demonstrate the vocalizer's ability and willingness to engage in physical confrontation, whereas submissive vocalizations show the contrastive pattern of high pitch and low amplitude.[4,45] Although the evidence for a relationship between body size and vocal pitch is mixed,[46,47] low-pitched voices are nonetheless interpreted by perceivers as belonging to larger individuals,[48] hence supporting the use of low pitch for aggressive vocalizations.

The aesthetic family of emotions was characterized by a complete reversal of both pitch and amplitude compared with the motivational emotions. Positive aesthetic vocalizations had both low pitch and low amplitude, whereas negative aesthetic vocalizations had high pitch and high amplitude. Although disgust is often included in studies of emotional vocalizations,[5,49] its positive counterpart—sensual pleasure—is very seldom studied, acoustically or otherwise. One study that reported vocal acoustics for this emotion[8] corroborated the findings here: vocal expressions of sensual pleasure had the lowest pitch of any emotion in that study.

The component process model

An alternative experimental design might have grouped emotions according to another cognitive appraisal theory of emotion, such as the well-known component process model.[50] However, such an approach would have grouped emotions in a manner very similar to that adopted in the present study and would have consequently reached similar conclusions.

Scherer[50] posited that vocal acoustics can show contrastive patterns across emotion classes and further provided clear and physiologically informed predictions for the five appraisal dimensions of the component process model. One of these dimensions is "intrinsic pleasantness," which is analogous to aesthetic appraisal in the OCC model. Disgust sits at the unpleasant end of this appraisal dimension, whereas pleasure sits at its positive end. Scherer[50] argued that because disgust involves rejecting offensive material from the mouth, its vocal expression may be influenced by the preparation to vomit. This proposal seems plausible in light of the fact that many of the muscles involved in airway protection during vomiting are also involved in pitch control. Indeed, the cricothyroid muscle—contraction of which raises vocal pitch—is tensed immediately before vomiting[51] but is relaxed before swallowing.[52,53] This leads to the prediction that vocal expressions of "intrinsically unpleasant emotions" will be high-pitched, whereas expressions of "intrinsically pleasant" emotions will be low-pitched. Laukka and Elfenbein[44] tested this prediction but did not find a relationship between intrinsic pleasantness and pitch. However, as discussed above, Scherer's[50] prediction was based on the physiology of disgust (and its opposite, pleasure), whereas Laukka and Elfenbein[44] evaluated intrinsic pleasantness with the questionnaire item "how pleasant was the event?" as applied across a broad set of emotions. The present study corroborated Scherer's[50] predictions in observing that the intrinsically unpleasant state of disgust was expressed with a higher pitch than the intrinsically pleasant state of pleasure.

Numinous emotions

"Numinous" emotions (awe and terror)—which are not present in the OCC framework or most other theories of emotion—were examined acoustically for the first time in this study. Keltner and Haidt[54] proposed that these emotions are experienced when an object is perceived as being so vast that it requires conceptual "accommodation" in order for it to be comprehended. Awe has been shown to have a recognizable vocal expression,[21] although no previous study has investigated the acoustics of vocally expressed awe or terror. Their vocal expressions in the present study did not follow the same trends as those of pleasure and disgust. There was no significant difference in pitch between awe and terror, although awe was significantly lower in amplitude than terror. The results suggested that these vocalizations may belong to a separate category of emotions. Given that experiences of awe and terror are described universally (although not exclusively) in religious contexts,[54,55] they may form a distinct "numinous" subtype of aesthetic emotions associated with the experience of deities and other stimuli perceived as being of large magnitude.

Rules of expression

Beyond simply showing that emotion families have distinct expression patterns, the larger goal of this study was to provide a new framework for thinking about vocal emotion through the formulation of general rules of expression, and this was nicely demonstrated by the observed interactions between acoustic patterns and emotion families. Figure 4 presents a framework for this analysis, which shows a 2 × 2 plot with Pitch on the *x*-axis and Loudness (as an intuitive proxy for amplitude) on the *y*-axis. Although this scheme is rooted in the OCC model and its trichotomy of emotion families, it looks beyond this model to a more general description of vocal acoustics.

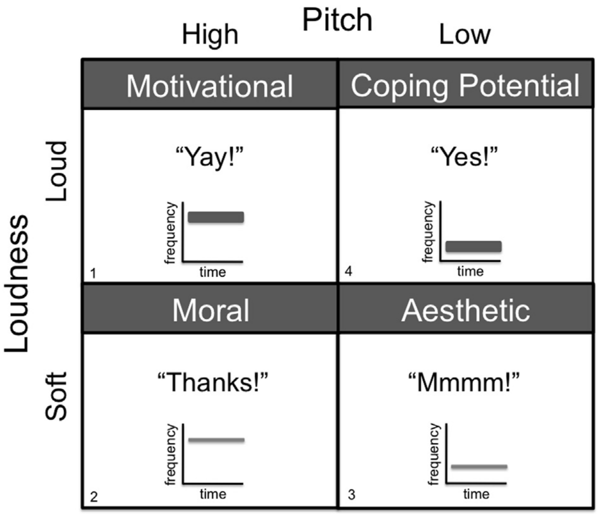

FIGURE 4. A framework for conceptualizing expression rules for vocal prosody. The figure presents a 2 × 2 plot of Loudness (y-axis) versus Pitch (x-axis), resulting in four possible expression patterns. (1) The "high-loud" combination is represented by the motivational emotions of the OCC model. (2) The "high-soft" combination is represented by the moral emotions. (3) The "low-soft" combination is represented by the aesthetic emotions. (4) Finally, the "low-loud" combination is not represented by any family in the OCC model but is compatible with an emotion family referred to in the literature as "coping potential," as exemplified by the exclamation of confidence, "Yes!". Note that the names for the expression rules reflect the profile for the *positive* member of a valence-pair for an emotion family. A schematic sonogram is shown below each exclamation. The height of the bar represents the relative pitch height of the exclamation, and the thickness of the bar represents its amplitude. These figures are not meant to be quantitative. Likewise, no attempt is made to represent pitch contours.

The cells in Figure 4 describe the acoustic profile of the *positive*-valenced emotion of a valence-pair. These four cells outline four possible expression rules for vocal prosody. In cell 1, positive emotions are higher in both pitch and amplitude than their negative counterparts, conforming to the "high-loud" rule (we use the psychological construct of "loudness" in these names, instead of the physical property of amplitude, to avoid the confusion of calling the rule "high-high"). The motivational emotions fit this profile. Positive emotions such as joy were higher in both pitch and amplitude than negative emotions such as distress. Cell 2 specifies positive emotions that are higher in pitch but lower in amplitude than negative counterparts, conforming to a "high-soft" rule. The moral emotions fit this profile. Positive emotions such as gratitude were higher in pitch but lower in amplitude than negative emotions such as anger. Cell 3 specifies positive emotions that are both lower in pitch and lower in amplitude than their negative counterparts, conforming to a "low-soft" rule. The aesthetic emotions fit this profile. Positive emotions such as pleasure were lower in both pitch and amplitude than negative emotions such as disgust. Finally, cell 4 specifies positive emotions that are lower in pitch but higher in amplitude than their negative counterparts, conforming to a "low-loud" rule. None of the three OCC categories fit this pattern. In looking to the literature on emotion, a good candidate for this expression rule is found in the appraisal dimension of "coping potential,"[50,56,57] where positive emotions with high coping potential, such as confidence, are spoken with a low pitch but high amplitude and negative emotions with low coping potential, such as nervousness, are spoken with a high pitch but low amplitude. As mentioned in the Introduction section, the presence of coping potential is one of the few points that distinguish the component process model from the OCC model. Overall, this plot accommodates the results obtained for the three emotion families examined in the present study and hypothesizes another acoustic valence-code related to coping potential as described by the component process model.[50]

Although the present study did not explicitly examine participants' ratings of affective intensity as a variable, previous work on affective prosody suggests other important expression rules, such as "intense-high-loud," where both pitch height and amplitude increase with affective intensity,[35,36] irrespective of valence. Hence, it is expected that even emotions that are uttered with low pitch would get higher (rather than lower) as affective intensity increases.

Limitations

It is important to point out the limitations of the results and of the 2 × 2 scheme. (1) The scheme is based on the idea that opposing valences within a given emotion family have opposing acoustic properties, which is what the present study found. However, other emotion classification schemes have different family structures or have no families at all. For example, within BET, happiness and sadness are not viewed as opposing emotions because that theory has no conception of emotion families but only discrete emotions. (2) There are emotion types that do not fit neatly within the scheme. As shown in the present data set, the so-called numinous emotions of awe and terror did not seem to fit any of the four expression rules. Similar exceptions will unquestionably be found for other emotions as well. (3) Perhaps most importantly, it would be wrong to view emotion families as homogeneous, either cognitively or acoustically. There are clearly *subtypes* within each family that the present study has not explored here, and these subtypes may actually conflict with the patterns that the study has described for the families. For example, although it is argued above that positive aesthetic emotions have low pitch ("Mmmm!"), there may be examples where this is not so. When people look at a baby or a puppy, they may be more likely to utter "Cute!" with a high pitch than a low pitch. Hence, there may be a basis for subtyping the Mmmm-type versus the Cute-type of aesthetic emotions. (4) Fear belongs to a subtype of motivational emotions, which would be predicted to have a low-soft expression pattern as a negative motivational emotion. In actuality, fear is expressed with high-loud vocalizations.[8] This pattern might be best explained by the high-affective intensity form of fear because affective intensity increases both pitch and amplitude. This discrepancy can be reconciled if "anxiety" is taken as the low-affective intensity form of fear. The vocal expression of anxiety does fit the profile of a motivational emotion under this scheme.[5] (5) These data were provided by undergraduates unselected for theatrical training. Although the vocalizations of professional actors contain systematic differences from spontaneous vocalizations for some acoustic parameters,[58] untrained speakers may be more variable in their performance.

CONCLUSIONS

The present study sought to examine whether the acoustic valence-code of vocal expressions of emotion is consistent across broad emotional families by analyzing unrehearsed exclamations produced by a relatively large sample of subjects. The results demonstrated a weak but significant effect of valence but a strong interaction with emotion family. The present study also provides the first acoustic description of vocal expressions for the so-called numinous emotions. The results were conceptualized as a 2 × 2 (Pitch × Loudness) scheme, setting up four possible rules of vocal expression. Each of the emotion families showed a unique placement in the cells of this scheme, with motivational emotions showing a "high-loud" pattern; moral emotions, a "high-soft" pattern; and aesthetic emotions, a "low-soft" pattern. The fourth cell of the scheme, representing the "low-loud" pattern, was not accounted for by any of the emotion families in the present study but was hypothesized to be consistent with emotions related to coping potential as described by the component process model.[50] These results demonstrate that the acoustic valence-code is not constant but varies as a function of emotion family. The inconsistency of reported acoustic valence-codes in the literature may be due to differences in the emotions considered by individual studies. Although further work in this area is clearly needed, the current Pitch × Loudness scheme provides an informative means of thinking about the acoustics of vocal expression and about emotion classification more generally.

REFERENCES

1. Juslin PN, Laukka P. Communication of emotions in vocal expression and music performance: different channels, same code? *Psychol Bull*. 2003; 129:770–814.
2. Scherer KR. Expression of emotion in voice and music. *J Voice*. 1995;9: 235–248.
3. Morton E. On the occurrence and significance of motivation-structural rules in some bird and mammal sounds. *Am Nat*. 1977;111:855–869.
4. Owings DH, Morton ES. *Animal Vocal Communication: A New Approach*. New York, NY: Cambridge University Press; 1998.
5. Banse R, Sherer K. Acoustic profiles in vocal emotion expression. *J Pers Soc Psychol*. 1996;70:614–636.
6. Lieberman P, Michaels SB. Some aspects of fundamental frequency and envelope amplitude as related to the emotional content of speech. *J Acoust Soc Am*. 1962;34:922–927.
7. Fairbanks G, Pronovost W. Vocal pitch during simulated emotion. *Science*. 1938;88:382–383.
8. Belin P, Fillion-Bilodeau S, Gosselin F. The Montreal affective voices: a validated set of nonverbal affect bursts for research on auditory affective processing. *Behav Res Methods*. 2008;40:531–539.
9. Schröder M. Experimental study of affect bursts. *Speech Commun*. 2003; 40:99–116.
10. Thompson WF, Balkwill LL. Decoding speech prosody in five languages. *Semiotica*. 2006;158:407–424.
11. Szameitat DP, Alter K, Szameitat AJ, Wildgruber D, Sterr A, Darwin CJ. Acoustic profiles of distinct emotional expressions in laughter. *J Acoust Soc Am*. 2009;126:354–366.
12. Fernald A, Simon T. Expanded intonation contours in mothers' speech to newborns. *Dev Psychol*. 1984;20:104–113.
13. Fernald A. Four-month-old infants prefer to listen to motherese. *Infant Behav Dev*. 1985;8:181–195.
14. Green JA, Whitney PG, Potegal M. Screaming, yelling, whining, and crying: categorical and intensity differences in vocal expressions of anger and sadness in children's tantrums. *Emotion*. 2011;11:1124–1133.
15. Bastian A, Schmidt S. Affect cues in vocalizations of the bat, Megaderma lyra, during agonistic interactions. *J Acoust Soc Am*. 2008;124:598–608.
16. Knutson B, Burgdorf J, Panksepp J. Ultrasonic vocalizations as indices of affective states in rats. *Psychol Bull*. 2002;128:961–977.
17. Gobl C, Chasaide AN. The role of voice quality in communicating emotion, mood and attitude. *Speech Commun*. 2003;40(1–2):189–212.
18. Hoffmann R, Mixdorff H (Eds.). Third International Speech Prosody Conference; 2006 May 2-5; Dresden, Germany. France: ISCA Archive, 2003.
19. Ekman P, Sorenson ER, Friesen WV. Pan-cultural elements in facial displays of emotion. *Science*. 1969;164:86–88.
20. Sauter DA, Eisner F, Calder AJ, Scott SK. Perceptual cues in nonverbal vocal expressions of emotion. *Q J Exp Psychol*. 2010;63:2251–2272.
21. Simon-Thomas ER, Keltner DJ, Sauter D, Sinicropi-Yao L, Abramson A. The voice conveys specific emotions: evidence from vocal burst displays. *Emotion*. 2009;9:838–846.
22. Russell JA. Core affect and the psychological construction of emotion. *Psychol Rev*. 2003;110:145–172.
23. Russell JA. A circumplex model of affect. *J Pers Soc Psychol*. 1980;39: 1161–1178.
24. Ortony A, Clore GL, Collins A. *The Cognitive Structure of Emotions*. New York, NY: Cambridge University Press; 1988.
25. Darwin C. *The Expression of the Emotions in Man and Animals*. 3rd ed. New York, NY: Oxford University; 1872.
26. Scherer KR, Wallbott HG. Evidence for universality and cultural variation of differential emotion response patterning. *J Pers Soc Psychol*. 1994;66: 310–328.
27. Titze IR. On the relation between subglottal pressure and fundamental frequency in phonation. *J Acoust Soc Am*. 1989;85:901–906.
28. Scherer K. Vocal communication of emotion: a review of research paradigms. *Speech Commun*. 2003;40(1–2):227–256.
29. Friberg A, Bresin R, Sundberg J. Overview of the KTH rule system for musical performance. *Adv Cogn Psychol*. 2006;2:145–161.
30. Gramming P, Sundberg J, Ternstrom S, Leanderson R, Perkins WH. Relationship between changes in voice pitch and loudness. *J Voice*. 1988;2:118–126.
31. Fonagy I. A new method of investigating the perception of prosodic features. *Lang Speech*. 1978;21:34–49.
32. Ilie G, Thompson WF. A comparison of acoustic cues in music and speech for three dimensions of affect. *Music Percept*. 2006;23:319–330.
33. Goudbeek M, Scherer K. Beyond arousal: valence and potency control cues in the vocal expression of emotion. *J Acoust Soc Am*. 2010;128:1322–1336.
34. Laukka P, Juslin P, Bresin R. A dimensional approach to vocal expression of emotion. *Cogn Emot*. 2005;19:633–653.
35. Bachorowski JA, Owren MJ. Vocal expression of emotion: acoustic properties of speech are associated with emotional intensity and context. *Psychol Sci*. 1995;6:219–224.
36. Juslin PN, Laukka P. Impact of intended emotion intensity on cue utilization and decoding accuracy in vocal expression of emotion. *Emotion*. 2001;1:381–412.
37. Boersma P, Weemink D. Praat, a system for doing phonetics by computer. *Glot Int*. 2011;5(9–10):341–345.
38. Wier CC, Jesteadt W, Green DM. Frequency discrimination as a function of frequency and sensation level. *J Acoust Soc Am*. 1977;61:178–184.
39. Whalen DH, Levitt AG. The universality of intrinsic F0 of vowels. *J Phon*. 1995;23:349–366.
40. R Core Development Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing; 2011.
41. Bates D, Maechler M. lme4: Linear Mixed-effects Models Using S4 Classes [software]. R Package Version 0.999375-37. 2010. http://cran.r-project.org.

42. Curtis V, Aunger R, Rabie T. Evidence that disgust evolved to protect from risk of disease. *Proc Biol Sci*. 2004;271(suppl):S131–S133.
43. Oaten M, Stevenson RJ, Case TI. Disgust as a disease-avoidance mechanism. *Psychol Bull*. 2009;135:303–321.
44. Laukka P, Elfenbein HA. Emotion appraisal dimensions can be inferred from vocal expressions. *Soc Psychol Personal Sci*. 2011;3:529–536.
45. Ohala J. An ethological perspective on common cross-language utilization of F0 of voice. *Phonetica*. 1984;41:1–16.
46. Evans S, Neave N, Wakelin D. Relationships between vocal characteristics and body size and shape in human males: an evolutionary explanation for a deep male voice. *Biol Psychol*. 2006;72:160–163.
47. Lass NJ. Correlational study of speakers' heights, weights, body surface areas, and speaking fundamental frequencies. *J Acoust Soc Am*. 1978;63: 1218–1220.
48. Feinberg D, Jones B, Little A, Burt D, Perrett D. Manipulations of fundamental and formant frequencies influence the attractiveness of human male voices. *Anim Behav*. 2005;69:561–568.
49. Scherer KR, Banse R, Wallbott HG, Goldbeck T. Vocal cues in emotion encoding and decoding. *Motiv Emot*. 1991;15:123–148.
50. Scherer KR. Vocal affect expression: a review and a model for future research. *Psychol Bull*. 1986;99:143–165.
51. Lang IM, Dana N, Medda BK, Shaker R. Mechanisms of airway protection during retching, vomiting, and swallowing. *Am J Physiol Gastrointest Liver Physiol*. 2002;283:G529–G536.
52. Doty RW, Bosma JF. An electromyographic analysis of reflex deglutition. *J Neurophysiol*. 1956;19:44–60.
53. Thexton AJ, Crompton AW, German RZ. Electromyographic activity during the reflex pharyngeal swallow in the pig: Doty and Bosma (1956) revisited. *J Appl Physiol*. 2007;102:587–600.
54. Keltner D, Haidt J. Approaching awe, a moral, spiritual, and aesthetic emotion. *Cogn Emot*. 2003;17:297–314.
55. Grandjean D, Rendu AC, MacNamee T, Scherer KR. The wrath of the gods: appraising the meaning of disaster. *Soc Sci Inf*. 2008;47:187–204.
56. Smith CA, Ellsworth PC. Patterns of cognitive appraisal in emotion. *J Pers Soc Psychol*. 1985;48:813–838.
57. Scherer KR. Appraisal considered as a process of multilevel sequential checking. In: Scherer KR, Schorr A, Johnstone T, eds. *Appraisal processes in emotion: theory, methods, research*. New York, NY: Oxford University; 2001:92–120.
58. Jürgens R, Hammerschmidt K, Fischer J. Authentic and play-acted vocal emotion expressions reveal acoustic differences. *Front Psychol*. 2011;2: 180–191.