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