journal of interdisciplinary music studies
spring/fall 2013, volume 7, issue 1&2, art. #13071202, pp. 29-44

Tone of voice in emotional expression and its implications for the affective character of musical mode

Daniel Bowling[1], Bruno Gingras[1], Shui'er Han[2], Janani Sundararajan[3], and Emma Opitz[4]

[1] Department of Cognitive Biology, University of Vienna
[2] School of Psychology, University of Sydney
[3] Department of Neurobiology, Duke University
[4] Cognitive Science, Hampshire College

Background in musicology. Although musical "mode" may be used to refer to a variety of different concepts, a simple definition is: a collection of tones and tone-relationships used to construct a melody. Using this definition, it is clear that associations between specific modes and emotions are formally documented and widely disseminated in a variety of musical traditions. Despite this rich cross-cultural history, the reasons why we perceive certain modes as best-suited for the expression of particular emotions remains unclear.
Background in Perception Science. Experience gathered over phylogenetic and ontogenetic time has prepared the human auditory system to rapidly associate various patterns of acoustic energy with their significance for behavior. This suggests that understanding our perception of tones in music requires understanding the biological significance of tones in the human auditory environment. Defining "tones" as sounds with regularly repeating structure that we perceive as having pitch, it seems reasonable to assume that the most biologically relevant and frequently experienced sources of tones is the human voice. This perspective implies that understanding mode-emotion associations in music begins with understanding the tonal properties of emotional expression in the voice.
Aims. To investigate the tonal properties of vocal emotional expression, examine their relationship to the perception of emotional arousal and valence, and discuss the implications for a biologically grounded theory of mode-emotion associations in music.
Main contribution. The tonal properties of the voice differ as a function of emotional arousal and valence, and are correlated with the perception of these qualities. Examination of tonal properties relevant to vocal expression in the context mode-emotion associations suggests that they also play a role in communicating emotion in music.
Implications. The tonal properties of vocal emotional expression can help us understand widespread associations between emotion and musical mode.

Keywords: Tone of voice, musical mode, emotion, arousal, valence.

•*Correspondence:* Daniel Bowling, Department of Cognitive Biology, University of Vienna.
tel: +431427776101; email: dan.bowling@univie.ac.at
• *Received:* 02 October 2013; *Revised:* 25 April 2014; *Accepted:* 19 May 2014
• *Available online:* 22 June 2014
• doi: 10.4407/jims.2014.06.002

Introduction

Over the last 80 years, the role of acoustical properties such as tempo and intensity in communicating emotion through music has received considerable attention. As a result we know a great deal about the relationship between various levels of such properties and emotional percepts they sponsor (reviewed in Gabrielsson and Juslin, 2003; and Juslin and Laukka, 2003). This literature stands in parallel to another large body of work focused on the acoustical properties of vocal emotional expression (reviewed in Scherer, 1986; and Banse and Scherer, 1996). In 2003, Juslin and Laukka conducted a comprehensive review of the musical and vocal expression literatures and found significant overlap in the acoustical properties used to express the same emotions in both domains (see also Ilie and Thompson, 2006). Some of the most prominent overlaps concerned tempo, intensity, high-frequency energy, fundamental frequency (F0) level, F0 variability, F0 contour, onset shape, and microstructural irregularity. Following Spencer (1857; 1890), Juslin and Laukka argued that acoustical overlap in musical and vocal emotional expression is best explained by the theory that music expresses emotion by imitating the voice, thus gaining access to neural mechanisms that evolved to generate appropriate responses to vocalizations.

Although there is wide acceptance for acoustical similarity in musical and vocal emotional expression, certain aspects of music have traditionally been excluded from comparison (Juslin and Laukka, 2003, p. 774). One prominent example is musical mode, which we define here as the collection of tones and tone-relationships used to create a melody[i]. Despite the fact that associations between specific modes and emotions are found in a variety of musical traditions (e.g., Western [Zarlino, 1558/1968], Indian [Chelladurai, 2010], Persian [Nettl, 1986], and Japanese [Hoshino, 1996]), and in many cases date to antiquity (Plato, trans. 1955; Capwell, 1986), relatively few authors have addressed the possibility that mode-emotion associations are also related to the acoustical properties of vocal emotional expression. One reason for this is that the concept of mode is not directly applicable to the voice. Whereas music typically consists of discrete pitches with well-defined F0 relationships, voice pitch fluctuates more-or-less continuously with no discrete or stable F0 values that would lend to mode-like organization (Patel, 2008). Although this fact prevents any direct comparison between modes and vocal expressions, there is evidence that the melodies derived from different modes exhibit characteristic pitch patterns, and that these patterns resemble those found in the vocal expression of different emotions. Specifically, melodies composed in modes associated with joyful emotion and speech expressing joy are characterized by a greater frequency of large changes in F0, whereas melodies composed in modes associated with sad emotion and speech expressing sadness both comprise a larger proportion of small changes in F0 (Bowling, Sundararajan, Han, and Purves, 2012). Like Juslin and Laukka (2003), several of the present authors (D.B., S.H., and J.S.) interpreted this acoustical similarity between modal melodies and vocal expression in support of the hypothesis that music conveys emotion in part by imitating the voice, thereby extending

Spencer's hypothesis to the affective character of musical modes (see also Curtis and Bharucha, 2010; Bowling, 2013).

One problem with this proposed explanation is that the extent of F0 changes in vocal expression is likely to primarily track differences in emotional arousal (with small F0 changes corresponding to low arousal, and large changes corresponding to high arousal; Banse and Scherer, 1996; Bachorowski, 1999; Juslin and Laukka, 2001), whereas musical modes are typically associated with emotions characterized by specific combinations of both emotional arousal and valence. For example, the Western major mode is typically associated with high-arousal positive-valence emotions (e.g., joy; Zarlino, 1558/1968; Hevner, 1935; Crowder, 1984; 1985; Gerardi and Gerken, 1995; Peretz, Gagnon, and Bouchard, 1998; Gagnon and Peretz, 2003), whereas the South Indian *varaali raga* and Persian *shur dastgah* are typically associated with low-arousal negative-valence emotions (e.g., sadness; Nettl, 1986; Chelladurai, 2010). If the extent of F0 changes in vocal expression only communicates differences in emotional arousal, imitation of this property alone cannot account for the specific arousal-valence combinations observed in mode-emotion associations.

There is no reason, however, to restrict the relationship between modal melodies and vocal expression to the extent of F0 changes. Vocal expression involves dozens of acoustical properties, many of which are specific to tones and thus also have the potential to contribute to the affective character of mode. Here, we investigate such "tonal" properties in speech conveying joy, anger, relaxation, and sadness to determine which, if any, distinguish between emotions that are similar in terms of arousal but different in terms of valence. We address this issue with respect to production and perception by examining tonal properties in relation to the emotions intended by speakers as well as the emotions perceived by listeners. Having identified several relevant properties, we briefly consider what role they might play in music. Finally, we discuss the results in the context of a vocal explanation for the affective character of musical mode as realized in melodies.

Methods

Speech recordings. In the interest of cross-cultural validity, recordings of emotional speech were collected from native speakers of English, Mandarin, and Tamil (10 speakers for each language, 5 male). The reading material consisted of 5 short sentences (6-9 syllables) with emotionally neutral content (e.g. "the leaves are changing color"). They were written in English and translated into Mandarin and Tamil by S.H. and J.S. Each speaker was assigned a single sentence according to their native language and instructed to read it with the intention of expressing each of four different emotions: joy, anger, relaxation, and sadness. These emotions were selected to include examples of positive and negative valence at high and low levels of emotional arousal (see Russell, 1980). Recordings were made in the presence of an experimenter. Given that we did not use professional actors in this study, the experimenter sometimes needed to encourage speakers to produce authentic

emotional expressions, particularly if the speaker felt nervous or embarrassed by the experimental circumstances. Such encouragement took the form of "you must convince the listener that you are angry" and/or “I’m not convinced”. Critically, the experimenter never spoke the sentences herself nor provided specific directions as to how they should be read. After both the speaker and the experimenter were satisfied that the intended emotion had been expressed to the best of the speaker’s ability (typically within 4-5 readings), they moved on to the next emotion. This procedure resulted in a total of 120 recordings, balanced such that each emotion was performed by 5 males and 5 females in each language (see Supplementary Audio files 1-4 for examples). The recordings were made in an audiometric room (Eckel CL-12A) at Duke-NUS Graduate Medical School in Singapore. Speech was recorded at 44.1 kHz (16-bit) using a condenser microphone (Audio Technica AT4049a) coupled to a solid-state digital recorder (Marantz PMD670). The recording level was adjusted to an appropriate level for each speaker.

Tonal properties. All assessments were based on F0 and intensity values calculated using Praat (version 5.2.26). All input parameters to the “To Pitch” and “To Intensity” functions were set to their default values except for pitch range, which was expanded from 75-500 Hz to 60-800 Hz to accommodate emotionally expressive speech. The use of a single set of parameters to process speech from multiple speakers inevitably leads to errors in the algorithmic assignment of F0 values (e.g., octave errors and misattribution of voicing). As a precaution against these errors, data points with F0 values more than ±2.5 standard deviations away from the mean F0 for a given recording were excluded from further analyses (this amounted to an average of 1.8% [SD =2.4%] of the data points in a given recording). This was deemed favorable to individually adjusting the parameters for each speaker or emotional condition because it ensured that F0 values were always calculated in the same way.

Six tonal properties were calculated from each recording. These properties can be grouped into three categories. First, we examined *F0 mean* and *F0 standard deviation*. These are among the most commonly examined properties in studies of vocal emotion and thus served as a basis for comparison of the present recordings with those used in previous research. Second, we examined change in F0 over time. This was done on two time scales: (1) the relatively slower changes that occur between voiced speech sounds (i.e., vowels and voiced consonants); and (2) the more rapid changes that occur within voiced speech sounds. Changes between voiced speech sounds (hereafter referred as *prosodic intervals*) were measured by calculating the frequency differences (in cents[ii]) between adjacent local intensity maxima (Bowling et al. 2012). The locations of these intensity maxima were determined using a custom Matlab (version R2009) script designed to locate positive-to-negative zero crossings in the first derivative of the intensity contour supplied by Praat. Only intensity maxima where F0 was defined (i.e., voiced portions of the signal) were used in the calculation of prosodic intervals. The distribution of prosodic interval sizes for each recording was summarized by a single number equal to the proportion of intervals smaller than 200 cents. This interval was chosen as a cut-off because, in music, it differentiates melodies composed in modes associated with low-arousal/negative and high-arousal/positive emotions, which typically comprise more

intervals smaller and larger than 200 cents respectively (Bowling et al. 2012). Changes within voiced speech sounds (hereafter referred to as *F0 perturbations*) were measured by calculating the percent deviation of each continuous F0 contour from a smoothed version of itself. Smoothing was accomplished by low-pass filtering each contour at 20 Hz with a third-order Butterworth filter (Juslin and Laukka, 2001). Third, we examined the distribution of energy as a function of frequency (i.e., the acoustic spectrum). Spectra were assessed in terms of: (1) *average voiced spectra*, which gives an idea of the distribution of energy in voiced speech sounds; and (2) *Harmonics-to-noise ratio* (*HNR*), which is a comparison of the proportion of energy contained in F0 and its harmonics compared to that not contained in F0 and its harmonics (Boersma, 1993). Prior to spectral calculations, the amplitude (root mean square) of each recording was normalized to the same value to ensure overall comparability in the amount of energy across recordings. Spectra were calculated using Matlab. 50 millisecond speech segments centered on voiced local intensity maxima were extracted from each recording, windowed, and Fourier transformed (see Matlab functions "hamming.m" and "fft.m"). The resulting spectra were averaged to obtain the average voiced spectrum for a given recording. This average spectrum was summarized by a single number equal to the energy above 750 Hz (calculated as a proportion of the total energy from 0-10000 Hz). 750 Hz was chosen as a cut-off because it is approximately where the pattern of spectral energy reverses for low-arousal and high-arousal expressions (see dashed line in Figure 3A). HNR was calculated using the Praat function "To Harmonicity". All input parameters to this function were set to their default values with the exception of pitch floor, which was maintained at 60 Hz for consistency with the "To Pitch" analysis (see above). Only HNR values at voiced local intensity maxima were included in further analyses.

The perception of emotion in the speech recordings. A separate group of 30 participants (mean age = 25, *SD* = 5; 15 male) was recruited at the University of Vienna to rate the emotion they perceived in the speech recordings. Most of these participants were native German speakers with varying degrees of English ability. Two spoke Mandarin, and none spoke Tamil. Ratings were made using a graphical user interface programmed in LiveCode (version 6.0.0). On each trial, the participant clicked a button to listen to a recording (repeatedly if necessary) and then rated the emotion they perceived by using the mouse to click somewhere within a two-dimensional "emotion space" with axes for arousal (running from low to high) and valence (running from negative to positive) that crossed in the center (a variant of Russell's [1980] circumplex model of emotion). Responses were recorded as x-y coordinates each running from -1 to +1 (resolution = 0.01). Each participant completed 132 trials in randomized order, these consisted of the 120 speech recordings, and 12 repetitions used to assess intra-rater reliability. The emotion space was explained in written instructions given to each participant at the beginning of the rating experiment. These instructions included example emotions for each quadrant that corresponded to the emotions intended by the speakers (i.e., high-arousal/positive-valence = joy, high-arousal/negative-valence = anger, low-arousal/positive-valence = relaxation, and low-arousal/negative-valence = sadness). After instruction, the participant's understanding of the emotion space was tested by

asking them to describe the emotion represented by several points randomly selected within the space.

The average ratings of arousal and valence (calculated across all 30 raters) were compared with the values of each tonal property using regression analyses. Because each speaker contributed four recordings to the data set (one in each emotional condition), "speaker" was included as a categorical covariate in these analyses following the method for calculating within-subject correlation coefficients with repeated observations described by Bland and Altman (1995). The resulting *r*-values describe the strength of the relationship between ratings of arousal or valence and each tonal property while accounting for variance due to individual speaker differences. These analyses were repeated 3 times: once including expressions for all four emotions examined; once including only low-arousal expressions (relaxation and sadness); and once including only high-arousal expressions (joy and anger). The latter two repetitions were important because if a given property was only relevant for distinguishing emotions at, for example, high-arousal levels, the effect might not be noticed when looking across all four emotions. Significance levels were Bonferroni corrected for all comparisons involving the same combination of predictor (i.e., tonal property) and dependent variable (i.e., ratings of arousal or valence).

Analysis of tones produced by musical instruments. Although all of the tonal properties described above can be measured in music, most are determined by the particulars of the melody under consideration (e.g., F0 mean, F0 standard deviation, the proportion of small/large intervals, and to some extent, the proportion of energy above 750 Hz). F0 perturbation and HNR, however, have more to do with the instrument making the music and how a performer uses it. Given that F0 perturbation and HNR were found to differentiate vocal expressions of different emotions (see Results), we assessed the potential role these properties play in music by measuring them in tones produced by traditional musical instruments. For this purpose, recordings of 17 different musical instruments (Alto Saxophone, Bass Clarinet, Bass Trombone, Double Bass, Clarinet, Cello, Flute, French Horn, Guitar, Oboe, Piano, Soprano Saxophone, Tenor Trombone, Trumpet Tuba, Viola, and Violin) were obtained from the University of Iowa's musical instrument samples database (Fritts, 1997). Each recording comprised between 8-12 tones with frequencies in the C3, C4, or C5 octaves (approximately 131-247 Hz, 262–494 Hz, or 523–988 Hz respectively) in accordance with the middle range of the instrument in question. It is important to note that these recordings were not produced with any particular emotional intentions and are therefore not directly comparable to the speech recordings. Accordingly, their analysis here serves only to provide an idea of the levels of F0 perturbation and HNR typically found in musical tones, which will be useful in understanding potential relations between emotion in voice and music (see Discussion).

The methods used to calculate F0 perturbation and HNR in the musical instrument recordings were the same as those used for the speech recordings with two exceptions, the pitch range of the "To Pitch" function was changed to 100–1100 Hz to accommodate the F0 values of the musical tones under consideration, and the pitch floor of the "To Harmonicity" function was maintained at 100 Hz for consistency.

The 50 millisecond windows used for HNR calculations were extracted from one of two possible locations depending on the shape of a musical tones' amplitude envelope. For tones with relatively long sustain phases (produced, e.g., by the woodwind, brass, and bowed string instruments in our sample), the window was placed at the center of the waveform, defined as halfway between time points where the intensity first rose above and fell below 5% of maximum. For tones with short/nonexistent sustain phases (guitar, piano, and plucked string instruments), the window was placed halfway between the time points representing maximum intensity and decay of this intensity by 50%.

Results

Tonal properties as a function of intended emotion. The results of the F0 mean and standard deviation analyses are shown in Figure 1.

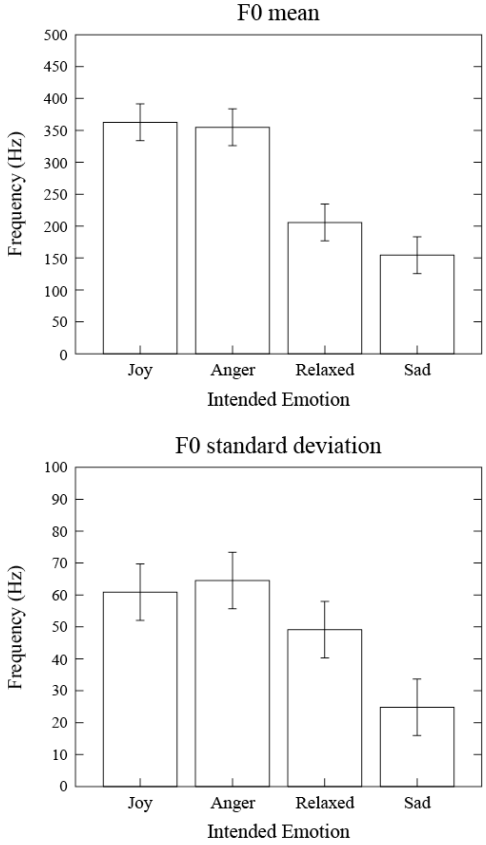

Figure 1. Average F0 mean and standard deviation in speech recordings as a function of intended emotion. Error bars indicate 95% confidence intervals.

Significant differences as a function of intended emotion were apparent for F0 mean, $F(3,87) = 195.81$, $p < 0.001$, and standard deviation, $F(3,87) = 14.21$, $p < 0.001$. In agreement with previous work, F0 mean was heavily influenced by the arousal level of the intended emotion (Banse and Scherer, 1996; Bachorowski, 1999; Juslin and Laukka, 2001), with significant differences apparent between joy and relaxation, joy and sadness, anger and relaxation, and anger and sadness (*ps* < 0.001; all p-values reported for pairwise comparisons in this section are Bonferroni-corrected). Although no significant difference was observed between high-arousal expressions (joy and anger), a significant difference was observed between low-arousal expressions (relaxation and sadness), with the mean F0 for sadness being lower than that for relaxation ($p < 0.001$). The results for F0 standard deviation were similar to those obtained for F0 mean, with the exception that relaxation exhibited more intermediate values between high-arousal expressions and sadness, and was only significantly different than sadness ($p < 0.005$). Thus, whereas neither F0 mean or standard deviation differed with valence between high-arousal expressions, they both differed with valence between low-arousal expressions. The results of the F0 change over time analyses are shown in Figure 2.

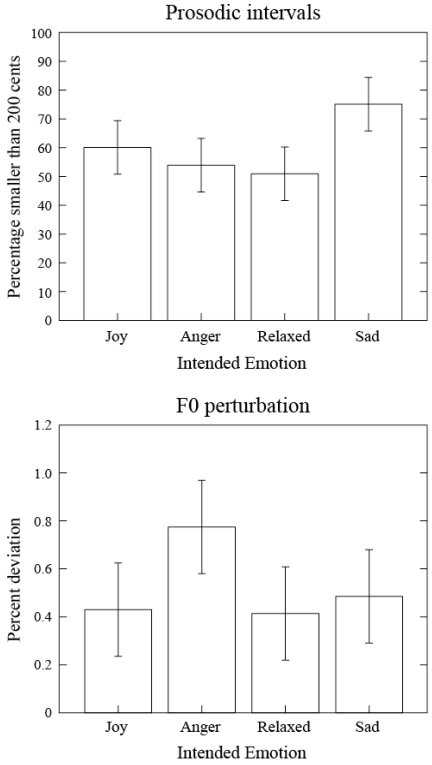

Figure 2. Average proportion of prosodic intervals smaller than 200 cents and average F0 perturbation in speech recordings as a function of intended emotion. Error bars indicate 95% confidence intervals.

Again, significant differences as a function of intended emotion were apparent for both prosodic intervals, $F(3,87) = 8.62$, $p < 0.001$, and F0 perturbation, $F(3,87) = 6.96$, $p < 0.005$. For prosodic intervals, this difference was driven by expressions of sadness, which comprised a greater proportion of small intervals than relaxation, joy, or anger ($ps < 0.05$). No other pairwise comparisons reached significance. For F0 perturbation, the only emotion that stood apart from the others was anger, for which perturbation values were nearly twice as high as those obtained for joy ($p < 0.05$) and relaxation ($p < 0.01$), and were also higher than values obtained for sadness, although this difference was not significant. Thus, whereas the proportion of small prosodic intervals differed with valence between low-arousal expressions (relaxation and sadness) but not high-arousal expressions (joy and anger), F0 perturbation differed with valence between high-arousal expressions but not low-arousal expressions.

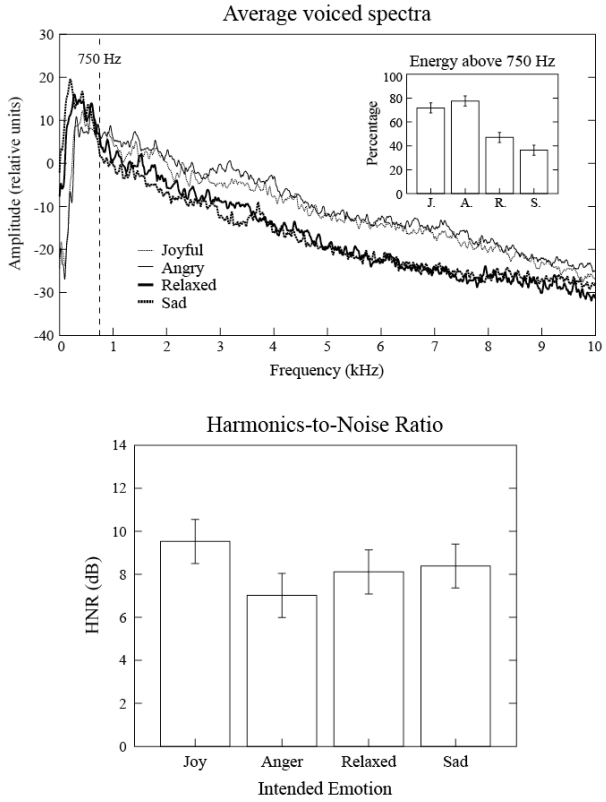

Figure 3. Average voiced spectra and harmonics-to-noise-ratio in speech recordings as a function of intended emotion. The inset in the top panel shows the average proportion of energy above 750 Hz (see Methods). Error bars indicate 95% confidence intervals.

The results of the spectral analyses are shown in Figure 3. Significant differences as a function of emotion were apparent for average voiced spectra, $F(3,87) = 132.83$, $p <$

0.001, and HNR, $F(3,87) = 3.89$, $p < 0.05$. For average voiced spectra, the pattern mirrored that observed for F0 mean in that the largest differences were between high- and low-arousal expressions ($ps < 0.001$). However, in contrast to F0 mean, where the only significant difference between similarly aroused emotions occurred between low-arousal expressions, significant differences in average voiced spectra (assessed in terms of the proportion of energy above 750 Hz; see inset) were observed between both relaxation and sadness ($p < 0.01$), as well as joy and anger ($p < 0.01$). However, these differences were relatively small (6-11%) compared to the differences observed between high- and low-arousal expressions (25-41%). For HNR, the only emotion that stood apart was joy, which was characterized by higher HNRs than all three of the other emotions, but only differed significantly from anger ($p < 0.01$). Thus, average voiced spectra differed with valence between both low- and high-arousal expressions, but HNR differed with valence between high-arousal expressions only.

Tonal properties as a function of perceived emotion. Agreement between raters (i.e., inter-rater reliability) was assessed by calculating the intra-class correlation (ICC; two-way random model variant) for ratings of arousal and valence across raters (Shrout and Fleiss, 1979). The single measures ICC was considerably higher for ratings of arousal (0.75, $p < 0.001$), than for ratings of valence (0.43, $p < 0.001$), indicating greater agreement between raters with respect to this response dimension. However, the average measures ICCs were very high for both ratings of arousal (0.99, $p < 0.001$) and valence (0.96, $p < 0.001$), justifying the use of average ratings (calculated across all 30 raters) in regression analyses with the tonal properties. The intra-rater reliability was assessed by comparing their responses to the first and second presentations of the 12 repeated stimuli using a variant of the ICC proposed by Eliasziw et al. (1994). The resulting values were 0.88 for arousal and 0.77 for valence ($ps < 0.01$), indicating "very high" and "high" reliability respectively (Landis, Koch, 1977).

The results of the regression analyses between tonal properties and ratings of arousal/valence are presented in Table 1. Across all four emotions, ratings of arousal increased with increases in F0 mean, F0 standard deviation, F0 perturbation, and the proportion of energy above 750 Hz, and decreased with increases in the proportion of small prosodic intervals (Table 1A, left). Looking only at low-arousal expressions (relaxation and sadness), the same relationships were observed for each tonal property except F0 perturbation, which was not significantly related to ratings of arousal in this subset (Table 1A, middle). Looking only at high-arousal expressions (joy and anger), ratings of arousal increased with F0 perturbation, the proportion of energy above 750 Hz, and HNR (Table 1A, right). Across all four emotions, ratings of valence decreased with increases in F0 perturbation; no other tonal properties were significantly correlated with ratings of valence overall (Table 1B). Looking only at low-arousal expressions, ratings of valence increased with increases in F0 mean, F0 standard deviation, and the proportion of energy above 750 Hz, and decreased with increases in the proportion of small prosodic intervals (Table 1B, middle). Looking only at high-arousal expressions, ratings of valence decreased with increases in F0 perturbation and the proportion of energy above 750 Hz, and increased with increases in HNR (Table 1B, right).

Table 1. Comparison of tonal properties and arousal/valence ratings

A Tonal Properties vs. Mean Arousal Ratings

Tonal property	Overall correaltion		Correlation for low-arousal only		Correlation for high-arousal only	
r	*r*	*p*	*r*	*p*	*r*	*p*
F0 mean	0.93	<0.0001*	0.79	<0.0001*	0.06	0.75
F0 SD	0.55	<0.0001*	0.55	0.0015*	0.11	0.12
pSmall PI	-0.29	0.0056*	-0.69	<0.0001*	0.2	0.29
F0p	0.26	0.013*	-0.2	0.29	0.7	<0.0001*
pAbove750	0.93	<0.0001*	0.81	<0.0001*	0.6	0.0003*
HNR	-0.13	0.24	-0.24	0.19	0.61	0.0003*

B Tonal Properties vs. Mean Valence Ratings

Tonal Property	Overall correlation		Correlation for low-arousal only		Correlation for high-arousal only	
r	*r*	*p*	*r*	*p*	*r*	*p*
F0 mean	-0.07	0.51	0.64	<0.0001*	0.13	0.49
F0 SD	0.07	0.51	0.52	0.0027*	0.083	0.66
pSmall PI	0.14	0.2	-0.76	<0.0001*	0.18	0.34
F0p	-0.36	0.0005*	-0.15	0.41	-0.45	0.012*
pAbove750	-0.17	0.11	0.64	0.0001*	-0.59	0.0004*
HNR	0.14	0.17	-0.24	0.19	0.55	0.0012*

F0m = F0 mean; F0 SD = F0 standard deviation; pSmall PI = Proportion of prosodic intervals smaller than 200 cents; F0p = F0 perturbation; pAbove750 = Proportion spectral energy above 750 Hz; HNR = Harmonics-to-noise ratio. (*Significant at Bonferroni corrected alpha level, 0.05/3 = 0.0167)

Finally, rater responses were also assessed as a function of the speaker's intended emotion category. This was done by examining how often rater responses fell within the appropriate quadrant of the emotion space for recordings of each intended emotion (Table 2). Inspection of Table 2 shows that all expressions were categorized at better-than-chance levels (assuming equal likelihood of responding in each quadrant, i.e., chance = 25%). In accordance with earlier work (e.g., Banse & Scherer 1996), anger was the most accurately recognized expression, followed by sadness, joy, and relaxation. Angry expressions were most frequently misinterpreted as joyful, joyful expressions as angry, relaxed expressions as joyous, and sad expressions as relaxed.

Table 2. Rater responses and intended emotions by category

Rated emotion	Intended emotion Joy	Anger	Relaxed	Sad
Joy	58 %	8.3 %	6.4 %	2.7 %
Anger	28.4 %	85.3 %	27 %	2.7 %
Relaxed	4.1 %	1.1 %	33.4 %	24.1 %
Sad	5.4 %	2.4 %	20.3 %	63 %

Columns do not add to 100% because a proportion of ratings for each intended emotion had 0 valence or 0 arousal (i.e., the rater clicked directly on the valence or arousal axis of the emotion space), and thus could not be categorized into one of the four quadrants.

Analysis of Musical Tones. Examination of F0 perturbation and HNR in tones produced by musical instruments showed that, across the instruments examined, musical tones were characterized by small perturbation values (M = 0.038%, SD = 0.038%) and high HNRs (*mean* = 31.54 dB, SD = 8.31 dB). These values are orders of magnitude lower and higher than those observed in speech. Averaged across all speech recordings, F0 perturbation was equal to 0.53% (SD = 0.38%), and HNR was equal to 8.26 dB, (SD = 3.35 dB). Thus, in comparison to speech, musical tones produced by traditional instruments tend to have very smooth F0 contours and be highly harmonic.

Discussion

The analysis of vocal emotional expression made here indicates that the tonal properties of the voice differ as a function of emotional arousal and valence, and are correlated with the perception of these qualities. Properties that differentiate emotional valence were found at both low and high levels of emotional arousal. When intended/perceived arousal was low, positive valence (relaxation) was characterized by higher and more variable F0s, less small prosodic intervals, and more energy above 750 Hz than negative valence (sadness). When intended/perceived arousal was high, positive valence (joy) was characterized by less F0 perturbation, less energy above 750 Hz, and higher HNRs than negative valence (anger). These results have the following implications for the theory that melodies composed in various modes convey emotion by imitating the extent of F0 changes in vocal expression.

For melodies that make frequent use of small F0 changes, such as those composed in minor modes or the *Varaali raga* (Bowling et al., 2012), the observed emotional association with low-arousal negative emotion (e.g., sadness) but not low-arousal positive emotion (e.g., relaxation) can be explained solely on the basis of imitation of F0 changes in vocal expression. Even though the size of prosodic intervals in vocal expression was tightly correlated with emotional arousal (as predicted; see Introduction), expressions of low-arousal negative emotion still comprised significantly more small intervals than expressions of low-arousal positive emotion. This suggests, that a preponderance of small F0 changes may uniquely convey low-arousal negative emotion in the voice. Accordingly, when a mode is such that the melodies derived from it make frequent use of small F0 changes, similarity to this quality of vocal expression alone provides a plausible basis for observed associations with low-arousal negative rather than positive emotion.

For melodies that make frequent use of large F0 changes, such as those composed in the major mode or *Mohanam raga* (Bowling et al., 2012), an explanation based on vocal expression is more complex. The proportion of large prosodic intervals (equal to the complement of the proportion of small prosodic intervals) was elevated in all high-arousal expressions. Differences in this proportion were not significant with respect to intended or perceived valence (see Figure 2A and Table 1B). This suggests that the frequent use of large F0 changes in musical melodies may convey heightened

arousal, but is unlikely to also convey information about valence. Accordingly, musical imitation of prosodic interval size alone appears insufficient to account for the associations with high-arousal *and* positive-valence that are observed with respect to modes like those mentioned above.

The present results, however, indicate two additional tonal properties that could work together with interval size to account for these associations. Vocal expressions of high-arousal positive emotion comprised significantly less F0 perturbation and higher HNRs than expressions of high-arousal negative emotion. The analysis of musical instruments shows that these qualities are greatly exaggerated in musical tones produced by traditional instruments. Together, these results raise the possibility that our emotional interpretation of traditional musical tones is inherently biased towards high-arousal and positive-valence. Thus, when a mode is such that the melodies derived from it make frequent use of large F0 changes *and* are comprised of smooth highly harmonic tones, similarity to these qualities of vocal expression provides a plausible basis for observed associations with high-arousal positive rather than negative emotion.

Given the proposed role of F0 perturbation and HNR in biasing our perception of typical musical tones, one might ask how the very same tones can also be used to convey negative emotion (e.g., anger and/or sadness)? At least two points are relevant here. First, emotional expression in music is served by many different acoustical properties, only a small subset of which were considered here. Even though a melody might use smooth harmonic tones, if it also uses mostly small intervals and is played with low intensity at a slow tempo, the combination of these factors is presumably sufficient to overcome the affective contributions of low perturbation and high HNR. It has been shown, for example, that the affective contribution of mode itself can be similarly overcome by other factors such as high intensity or fast tempo (Heinlein, 1928; Gagnon and Peretz, 2003). Second, F0 perturbation and HNR, like other acoustical properties, are subject to variation in actual music performance. The extent of this variation depends on the instrument, but in many cases can be considerable. For example, large increases in F0 perturbation and decreases in HNR have been documented in recordings of violinists and electric guitarists attempting to convey sadness and anger in simple melodies (Gabrielsson and Juslin, 1996).

As a final point, it is necessary to briefly consider the role that extra-acoustic factors, such as experience in a particular culture, might play in determining the nature of mode-emotion associations (for a fuller treatment of this issue see Bowling, 2013). The reason this must be considered here is that most modern accounts of music and emotion hold that mode-emotion associations are learned through exposure to the conventions of a particular musical tradition (e.g., see Lundin, 1967; Trainor and Corrigall, 2010). For example, it has been demonstrated that mode-emotion associations develop with experience and strengthen over time (Gerardi and Gerken, 1996; Dalla Bella, Peretz, Rousseau, Gosselin, 2001). If this perspective is correct, then there is no vocal basis for mode-emotion associations. There are, however, several reasons to believe this perspective is misguided. First, it is important to note that evidence for learning does not imply absence of biological preparation (Marler,

1991). Second, when melodies composed in modes associated with similar emotions in different cultures are compared, the same pitch patterns are apparent (Bowling et al., 2012). Third, the similarities between modal melodies and tonal properties of vocal expression presented here fit into the much broader pattern of acoustic overlap between musical and vocal expression (Juslin and Laukka, 2003). Taken together, these reasons imply that an explanation for the affective character of mode as realized in melody must also take into account the biology of emotional expression. Here, we have offered one such explanation based on the tonal properties of vocal emotional expression. In conclusion, we argue that the evidence provided here supports the theory that the affective character of musical modes is best understood in terms of the physical characteristics and biological purposes of vocalization.

Acknowledgements

This work was supported by a grant from the National Science Foundation [BCS-0924181] awarded to Dale Purves, and a grant from European Research Council [No. 230604] awarded to Tecumseh Fitch. We would also like to thank Dale Purves and Brian Monson for comments on the manuscript.

References

Bachorowski, J. (1999). Vocal expression and perception of emotion. *Current Directions in Psychological Science*, *8,* 53-57.

Banse, R., & Scherer, K.R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology, 40(3),* 614-636.

Balkwill, L., & Thompson, W.F. (1999). A cross-cultural investigation of the perception of emotion in music: psychophysical and cultural cues. *Music Perception, 17,* 43–64.

Bland, J.M., & Altman, D.G. (1995). Calculating correlation coefficients with repeated observations: Part 1-correlation within subjects. *BMJ, 310*, 446.

Boersma, P. (1993). Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. *IFA Proceedings, 17,* 97-110.

Bowling, D.L., Sundararajan, J., Han, S., Purves, D. (2012). Expression of emotion in eastern and western music mirrors vocalization. *PLoS ONE*, *7(3)*, e31942.

Bowling, D.L. (2013). A vocal basis for the affective character of musical mode in melody. *Frontiers in Psychology, 4,* 464.

Capwell, C. (1986). South Asia. In A.M. Randel (Ed.), *The New Harvard Dictionary of Music* (pp. 778-787). Cambridge, MA: Belknap Press.

Chelladurai, P.T. (2010) *The splendor of South Indian music*. Dindigul, IN: Vaigarai Publishers. pp 31-32, 55-56, 101-102, 155-166, 129-130, 177.

Crowder, R.G. (1984). Perception of the major/minor distinction: I. Historical and theoretical foundations. *Psychomusicology, 4*, 3–12.

Crowder, R.G. (1985). Perception of the major/minor distinction: II. Experimental investigations. *Psychomusicology, 5*, 3–14.

Curtis, M.E., & Bharucha, J.J. (2010). The minor third communicates sadness in speech, mirroring its use in music. *Emotion, 10*, 335–48.

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,* B1-B10.

Eliasziw, M., Young, S.L., Woodbury, M.G., Fryday-Field, K. (1994) Statistical methodology for the concurrent assessment of interrater and intrarater reliability: using goniometric measurements as an example. *Physical Therapy, 74,* 777-788.

Fritts, L. (1997). *University of Iowa electronic music studios musical instrument samples*. [Data set]. Retrieved from http://theremin.music.uiowa.edu/MIS.html

Gabrielsson, A., Juslin, P.N. (1996). Emotional expression in music performance: between the performer's intention and the listener's experience. *Psychology of Music, 24,* 68-91.

Gabrielsson, A., Juslin, P.N. (2003). Emotional expression in music. In R.J. Davidson, K.R. Scherer, & H.H. Goldsmith (Eds.), *Handbook of affective sciences (*pp. 503-534). New York, NY: Oxford University Press.

Gagnon, L., & Peretz I. (2003). Mode and tempo relative contributions to "happy-sad'' judgements in equitone melodies. *Cognition & Emotion*, *17*, 25–40.

Gerardi, G.M., & Gerken, L. (1995) The development of affective responses to modality and melodic contour. *Music Perception, 12,* 279–290.

Hevner, K. (1935). The affective character of the major and minor modes in music. *The American Journal of Psychology, 47*, 103–118.

Heinlein, C.P. (1928). The affective character of the major and minor modes in music. *Journal of comparative psychology, 8,* 101-104.

Hoshino, E. (1996). The feeling of musical mode and its emotional character in a melody. *Psychology of Music, 24*, 29–46.

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

Landis, J.R., Koch, G.G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*, 159-174.

Lundin, R.W. (1967). *An objective psychology of music, 2nd edition*. New York, NY: The Ronald Press Company. pp. 166-169.

LiveCode. (Version 6.0.0) [Computer software]. Ediburgh, SCT: RunRev Ltd.

Marler, P. (1991). The instinct to learn. In S. Carey, and R. Gelman (Eds.), *The epigenesis of mind, essays on biology and cognition* (pp. 212-231). Hillsdale, NJ: Lawrence Erlbaum Associates.

Matlab. (Version R2009) [Computer software]. Natick, MA: MathWorks Inc.

Nettl, B. (1986). Near and Middle East. In A.M. Randel (Ed.), *The New Harvard Dictionary of Music* (pp. 528-524). Cambridge, MA: Belknap Press.

Patel, A.D. (2008). Music, language, and the brain. New York, NY: Oxford University Press. p. 183.

Peretz, I., Gagnon, L., & Bouchard, B. (1998). Music and emotion: perceptual determinants, immediacy and isolation after brain damage. *Cognition, 68,* 111-141.

Plato. (1955). *The Republic*. (D. Lee, Trans.). 2nd Edn. London, UK: Penguin Classics. 93-95. (Original work published ~375 BCE).

Praat: Doing Phonetics by Computer (Version 5.2.26) [Computer software]. Amsterdam, NL: Boersma, P., & Weenink, D.

Randel, D.M. (1986). Mode. In A.M. Randel (Ed.), *The New Harvard Dictionary of Music* (pp. 499-502). Cambridge, MA: Belknap Press.

Russell, J.A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology, 39,* 1161-1178.

Scherer, K.R. (1986). Vocal affect expression: a review and a model for future research. *Psychological Bulletin, 99,* 143-165.

Spencer, H. (1857). The origin and function of music. *Fraser's Magazine*, October, 396-408.

Spencer, H. (1890). Postscript to: the origin and function of music. *Mind*, os-XV.

Shrout, P.E., & Fleiss, J.L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin, 86,* 420-428.

Trainor, L.J., and Corrigall, K.A. (2010). Music acquisition and effects of musical experience. In M.R. Jones, A.N., Popper, and R.R. Fay (Eds.), *Music Perception (Springer Handbook of Auditory Research (*pp. 89-127). New York, NY: Springer.

Zarlino, G. (1968) *The art of counterpoint: part III of le institutioni harmoniche*. (G.A. Marco & C.V. Palisca, Trans.). New York, NY: The Norton Library. P. 21-22. (Original work published 1558).

---

[i] The term "mode" is used to refer to a variety of different concepts, only loosely related by their usage in the study of scales and melodies (Randel, 1986, 499). This ambiguity has generated considerable confusion, particularly when comparing different musical traditions. By avoiding additional related concepts such as rules for ornamentation or pre-specified melodic motifs, the simple definition provided here can readily be applied to music from a wide variety of traditions.

[ii] The cent is a logrithmic unit of frequency interval size. One cent is 1/100 of an equally-tempered semitone. An octave thus comprises 1200 cents. The formula for calculating the size of an interval between two frequencies (F1 and F2) in cents C, is $C = 1200*\mathrm{Log}_2(F1/F2)$.

Biographies

Daniel Bowling completed undergraduate degrees in psychology and neurophilosophy at the University of California at San Diego in 2006, and completed his PhD in neurobiology at Duke University in 2012. He is currently a postdoctoral fellow in the Department of Cognitive Biology at the University of Vienna. Broadly, he is interested in the natural history of nervous systems and the roles of selection and plasticity in shaping how we interact with the world. To date, these interests have manifested in research on the role of phylogenetic and ontogenetic experience with tonal sound stimuli in determining how we perceive tones in music.

Bruno Gingras completed an undergraduate degree in Biochemistry, a masters degree in Molecular Biology at Université de Montréal, and a PhD in music theory at McGill in 2008. His doctoral dissertation focused on expressive strategies and performer-listener communication in organ performance. From there he spent two years as a postdoctoral fellow at Goldsmiths where he continued his work on communication in music performance. His research interests include biomusicology, music performance, and the perception of musical structure.

Shui'er Han completed undergraduate in psychology at the National University of Singapore in 2010, after which she spent some time studying music and auditory perception with Dale Purves at Duke-NUS Graduate Medical School. She is currently a graduate student in the Department of Psychology at the University of Sydney studying cross-modal perception. She is broadly interested in perception and underlying neural mechanisms of cross-modal interactions.

Janani Sundararajan completed her undergraduate degree in Bioengineering at the National University of Singapore in 2010, after which she spent three years studying auditory and visual perception With Dale Purves at Duke-NUS Graduate Medical School. She is currently a graduate student in the Department of Neurobiology at Duke University and is broadly interested in understanding the neural circuits underlying perception and behavior.

Emma Optiz is an undergraduate student at Hampshire College in Amherst Massachusetts. Broadly she is interesting in the evolutionary origins of acoustic communication and music.