Emotion
2010, Vol. 10, No. 3, 335–348

1528-3542/10/$12.00 DOI: 10.1037/a0017928

The Minor Third Communicates Sadness in Speech, Mirroring Its Use in Music

Meagan E. Curtis and Jamshed J. Bharucha
Tufts University

There is a long history of attempts to explain why music is perceived as expressing emotion. The relationship between pitches serves as an important cue for conveying emotion in music. The musical interval referred to as the *minor third* is generally thought to convey sadness. We reveal that the minor third also occurs in the pitch contour of speech conveying sadness. Bisyllabic speech samples conveying four emotions were recorded by 9 actresses. Acoustic analyses revealed that the relationship between the 2 salient pitches of the sad speech samples tended to approximate a minor third. Participants rated the speech samples for perceived emotion, and the use of numerous acoustic parameters as cues for emotional identification was modeled using regression analysis. The minor third was the most reliable cue for identifying sadness. Additional participants rated musical intervals for emotion, and their ratings verified the historical association between the musical minor third and sadness. These findings support the theory that human vocal expressions and music share an acoustic code for communicating sadness.

Keywords: emotions, communication, prosody, music and emotion, musical intervals

The connection between music and emotion has puzzled humans for thousands of years. Circa 350 B.C., Aristotle wrote of the affective states induced by different musical modes (Jowett, 1885). Modern listeners use mode as a cue by which to differentiate between happy and sad music, but our understanding of the origin of these affective associations is still in its infancy.

We examined the possibility that pitch contours of emotional speech contain some of the same categorical pitch changes that convey emotion in Western tonal music. Others have suggested that human vocal expressions of emotion have served as a mapping source for music (Juslin, 1997; Juslin, 2001; Juslin & Laukka, 2003; Kivy, 1989; Papoušek, 1996; Scherer, 1995), and numerous correspondences have been noted across domains in the use of certain acoustic variables, such as mean fundamental frequency, intensity, and tempo (Juslin & Laukka, 2003).

Despite various correspondences, there is one looming disparity across domains. Music makes use of specific pitch relationships to convey emotion, whereas it has generally been assumed that speech does not. In music, these relationships are called *intervals* and correspond to the ratios between fundamental frequencies. The interval of a minor third is perceived to convey sadness (Cooke, 1959; Maher & Berlyne, 1982), whereas the major third is perceived to convey positive affect (Cooke, 1959). The majority of Western music is written in either the major or minor mode; these are associated with happiness and sadness, respectively (Crowder, 1985; Gerardi & Gerken, 1995; Hevner, 1935a, 1935b, 1936, 1937; Krumhansl, 1997; Nielzén & Cesarec, 1982; Peretz, Gagnon, & Bouchard, 1998; Wedin, 1972). The most salient difference between these modes is that the major mode is characterized by the interval of a major third between the third and first scale degrees (a frequency ratio of 5:4), whereas the minor mode is characterized by the interval of a minor third between these scale degrees (a frequency ratio of 6:5). Despite the general consensus that these intervals are involved in the musical communication of emotion, theories suggesting that the vocal expression of emotion serves as a mapping source for the acoustic parameters that communicate emotion in music (Juslin, 2001) do not account for the emotional expressiveness of musical intervals. This omission is curious, because speech seems to offer a clear mapping source for many other acoustic properties that communicate emotion in music (Juslin & Laukka, 2003). The modulation of pitch in speech may have an emotionally communicative function that has been largely overlooked by previous research.

The following experiments have been designed to examine whether the pitch contours of affective speech exhibit patterns similar to those used in the musical communication of emotion. In Experiment 1, actors were asked to recite bisyllabic utterances with four different emotions. These utterances were analyzed to determine whether pitch patterns from one syllable to the next varied systematically with emotion. Other acoustic parameters were also assessed. In Experiment 2, participants rated the recorded utterances for perceived emotion. Their ratings were used to model the use of acoustic parameters in decoding emotion. Experiment 3 assessed the emotional perception of musical intervals, so as to determine the cross-domain correspondences between intervals and emotion. Experiment 4 tested the perceptual categorization of intervals and was conducted to obtain empirical support for the interval categorizations utilized in the analyses of Experiments 1–3.

Meagan E. Curtis and Jamshed J. Bharucha, Department of Psychology, Tufts University.

We thank Kaivon Paroo and Gena Gorlin for research assistance. We also thank George Wolford, Jay Hull, Howard Hughes, Joanna Morris, and Neil Stillings for their guidance and valuable feedback.

Correspondence concerning this article should be addressed to Meagan E. Curtis, Department of Psychology, 490 Boston Avenue, Tufts University, Medford, MA 02155. E-mail: meagan.curtis@tufts.edu

Experiment 1

Utterances spoken by actors were analyzed to discern the pitch intervals from one syllable to the next. Each of four utterances was spoken with four different emotions: anger, happiness, pleasantness, and sadness. By using the same utterances across emotions, we controlled the semantic and syntactic content, minimizing prosodic variation due to factors other than the intended emotion. Each vocalization was bisyllabic, enabling us to measure the interval from the first syllable to the second.

Method

Participants. Nine female actresses were recruited from undergraduate theater groups at Tufts University. Previous research has shown that the vocal expressions of emotion produced by actors are similar to vocal expressions produced by nonactors in real, emotionally charged situations (Williams & Stevens, 1972). Each participant had several years of acting experience, with various levels of formal acting training exhibited across the group. The mean age of the participants was 19.3 years. Each actress was paid $10 for her participation.

Eight participants were raised in the United States, speaking American English as their first language. Two of these American actresses were raised in bilingual households and acquired English simultaneously with another language (Spanish and Korean). One actress was raised in Europe and had lived in London and Zurich. English was her first and only language. Although she was raised in Europe, she did not speak with a British or European accent. At the time of data collection, she had been living in the United States for 3 years and her accent was Americanized.

Stimuli. The stimuli comprised 16 emotionally charged scripted scenarios, along with the lines that each actress was to perform in response to each scenario. The four lines were "Let's go," "Okay," "Come here," and "Come on." Each line was paired with four different emotional scenarios, so that each line was recited with anger, happiness, pleasantness, and sadness. These four emotions were selected because they map onto different dimensions of valence and arousal (which we hoped would maximize the acoustic differences between emotions) and because they are among the most common expressive states (which should lead to accurate encoding and decoding).

Apparatus. The speech samples were recorded using a Boss BR-8 Digital Recorder and a Shure SM-58 microphone. The recordings were imported into Sound Forge 9.0 (Sony), a digital audio editing program, at a sampling rate of 44.1 kHz.

Procedure. Before participation in the study, the procedure was fully explained to each actress, and each gave written consent to participate and to have her voice recorded. Participants were asked to vocally communicate four emotions (happiness, sadness, anger, and pleasantness). They were given a script containing emotionally charged scenarios. They were asked to perform one scripted line in response to each scenario. Each participant was shown the script before her acting session. Each actress was informed that the recordings of her vocalizations would be used in future experiments and that her goal was to ensure that people would be able to identify the emotion conveyed in each vocalization. The actresses were instructed to make facial expressions corresponding to each specified emotion during their vocalizations. The act of making a facial expression can help induce the corresponding emotion (Ekman, Levenson, & Friesen, 1983), which should have helped the actresses convey each emotion.

The vocalizations were recorded in a small sound-attenuated room. Participants were given a microphone, which was connected to a digital recording device located outside of the room. Each actress was left alone in the room to record her vocalizations. Each actress read the script and performed the lines as though she were actually responding to the context established in the script. Participants were encouraged to repeat the vocalization for each scenario as many times as they wanted, allowing them to produce a genuine vocalization that they considered to be an adequate expression of the emotion. They were instructed that their final vocalization for each scenario would be used as their intended emotional portrayal for that particular scenario in the data analysis, unless they verbally indicated that they preferred a previous portrayal.

Acoustic analysis. Each actress recorded 16 bisyllabic vocalizations.[1] A syllable generally has one perceptually salient pitch. Thus, the pitch contour of each bisyllabic sentence contained two perceptually salient pitches, corresponding to one interval.

A total of 144 vocalizations were analyzed (four per emotion from each of the 9 participants) using Praat (Boersma, 2001) and the prosogram computational model (Mertens, 2004). Various acoustic analyses were conducted, including assessments of the pitch contour of each speech sample, which were the main analyses of interest. We also performed routine analyses, including the mean intensity, mean fundamental frequency, and duration of each speech sample.

The prosogram computational model (Mertens, 2004), which has been adopted by others studying prosodic pitch patterns (Patel, 2005; Patel, Iverson, & Rosenberg, 2006), was used to analyze the pitch contours of the speech samples. Prosogram recognizes that the fundamental frequency (F0) contour of speech is a physical measure and is not the most accurate measure of the perceived pitch contour of speech. Perceptually, the F0 contour is parsed into syllabic units, as indicated by rapid fluctuations in acoustic parameters (House, 1990). The brain calculates a time-weighted average of the F0 contour within the syllable, yielding the perceived pitch (D'Allesandro & Castellengo, 1994). Prosogram utilizes user-provided phonetic segmentation and calculates a time-weighted average of F0 within vowels. The analysis yields a pitch contour in discrete pitch units measured in hertz (Hz). The relationship between sequential pitches, which is referred to throughout this article as an interval, is calculated in cents, 1/100th of a musical semitone.

The pitch contour analysis was restricted to the portions of the speech samples corresponding to vowels and voiced consonants, the portions of the speech signal for which pitch can be calculated.[2]

[1] Examples of speech samples are available at http://ase.tufts.edu/psychology/music-cognition/emotion2009.html

[2] Although the prosogram model specifically computes a time-weighted average F0 for each vowel, we found it necessary to include voiced consonants as well as vowels for some speech samples. This was necessary when assessing a subset of the speech samples containing the word "come," as some speakers virtually omitted the "o" and instead vocalized the "m" with a clear, steady-state pitch. In instances in which the pitched portion of a syllable was voiced primarily as a consonant, the voiced consonant was included in the analysis. Otherwise, the pitch contour analysis focused on vowels.

Manual phonemic transcription identified the onset and offset of phonemes for the pitch contour analysis.

To analyze a speech sample, we set the analysis parameters in accordance with the F0 range of adult female speech. A F0 calculation range of 100 to 800 Hz was used, with a frame period of 0.005 s. The F0 contour and the intensity contour of each speech sample were calculated using the Praat autocorrelation method (Boersma, 1993).

Voicing (whether a frequency sample is voiced or unvoiced) was determined according to a default voicing threshold of 0.45, which is the strength of the unvoiced candidate relative to the maximum possible autocorrelation.

The calculations of F0, intensity, and voicing were used to segment the speech sample into vocalic nuclei. A vocalic nucleus is the portion of a voiced phoneme for which pitch is calculated, and this nucleus is determined in relation to the peak intensity that occurs in a voiced phoneme. Assessing the nucleus involves identifying the voiced portion of each phoneme that has sufficient intensity, which is determined by difference thresholds relative to the peak intensity of the phoneme.

The vocalic nuclei were then assessed according to a glissando threshold. This parameter is intended to model whether the vocalic nucleus would be perceived by human listeners as a steady-state pitch or as a glide (a pitch that rises or falls over its duration). The glissando threshold is the auditory threshold for pitch variation and is dependent on the degree and duration of frequency variation. Our analysis used a glissando threshold of $G = 0.32/T^2$, where T is the duration of the vocalic nucleus. This threshold was chosen because it has previously been shown to have a high degree of correspondence to the assessments made by auditory transcribers (Mertens, 2004).

The acoustic analysis yields the starting and ending times of the vocalic nuclei and a frequency value for each starting and ending point. These frequency values correspond to the pitch contour and are closely aligned with the F0 contour, as illustrated in Figure 1. The resulting pitch contour was then compared with the F0 contour of the speech sample. This enabled assessment of the accuracy of the pitch contour analysis. For some speech samples, a steady-state pitch was visible in the F0 contour and perceptible in the speech sample but was not detected by the Prosogram analysis. This was common when part of the phoneme corresponded to a glide. In such instances, the phonemic transcription was adjusted so that the portions of the voiced phoneme that corresponded to different pitch levels were treated as individual phonemes. Rapid spectral changes were used to guide the adjustments to the phonemic transcriptions, as rapid spectral changes are perceptual cues by which to parse F0 contours (House, 1990). The adjustments resulted in more precise and detailed analyses of the pitch contours. We also found it necessary to correct a small number of obvious octave errors that were present in the analyses, as F0 analysis is susceptible to such errors.[3]

Two salient, steady-state pitches were identified for every speech sample with the exception of two.[4] These steady-state pitches constitute the patterns of interest in the present study. However, we want to acknowledge that we have not attempted to quantify pitch glides in this investigation, which are a typical feature of the pitch contours in speech. Glides often occur during vocal transitions between steady-state pitches but also occur in instances that have no transitional properties. For instance, one's voice may glide downward or upward after sustaining a steady-state pitch, and such glides often have no perceptually salient ending point. Although our analyses focus on the steady-state pitches found in prosodic contours, we believe that glides are also communicative features of the acoustic signal that exhibit emotionally specific patterns. This assertion is based on our observations of glide patterns present in the current data set.[5] Because of the problematic nature of quantifying glide patterns, we have restricted our analyses to the steady-state pitch patterns of emotional speech.

The intervals between adjacent steady-state pitches, corresponding to f_1 and f_2, were computed in cents as follows:

$$\text{cents} = 1{,}200\ [\log(f_1/f_2)/\log 2].$$

It is necessary to use cents to compute the size of the frequency change between adjacent pitches, because the Hz is insufficient for capturing perceptual pitch relationships. Pitch perception functions logarithmically with respect to frequency, so it is necessary to transform frequency relationships according to a scale that captures pitch perception in a linear manner. Cents, a unit of measurement along the musical semitone scale, does just that.

Results

Figure 2 illustrates the distribution of speech intervals according to the intended emotion of the speaker. Musical intervals are perceived as belonging to semitone categories separated by roughly 100 cents, with a sharp perceptual discrimination boundary falling at 50 cents between musical interval category prototypes (Burns & Campbell, 1994; Burns & Ward, 1978; Siegel & Siegel, 1977a, 1977b). Consequently, intervals were coded categorically in 100-cent bins. Six intervals were larger than once octave, and these intervals were normalized to within octave range for all categorical analyses. Normalization through octave equivalence is standard practice in the conceptualization of pitch class hierarchies in tonal music.

We considered the band of the distribution within which most of the intervals occurred. This is the band within which differentia-

[3] Octave errors were only corrected when they were an obvious feature of the visualized F0 contour. For instance, the bell curve shape is a common characteristic of visualized F0 contours. An unnatural disruption of the bell curve shape served as an obvious indicator of an octave error. Such errors were verified perceptually and corrected.

[4] One speech sample was whispered and had no F0 contour (and was thus omitted from F0 analyses). Another speech sample contained a steady-state pitch in the first syllable and a long, slow glide in the second syllable. The glide spanned approximately 1 semitone from start to finish but contained a pitch that was identifiable through direct perception. An independent auditory transcriber perceptually determined the pitch of this syllable, and this pitch assessment was used for data analysis.

[5] We assessed the ratio of the duration of steady-state pitches to the duration of frequency glides across all speech samples. The grand mean ratio of pitch duration to glide duration was 1.07, indicating that the speech samples as a whole comprised equal proportions of glides and steady-state pitches. Differences in the proportions of pitch duration to glide duration were observed between the four emotions (glides were more prevalent than steady-state pitches in the happy and angry speech samples, and the opposite pattern was observed for the pleasant and sad speech samples), but these differences did not reach the level of significance.

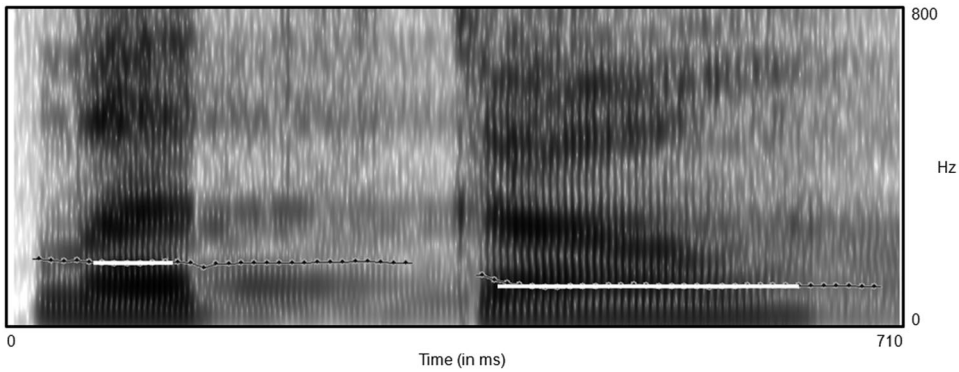

Figure 1. Spectrogram of a speech sample. The pitch contour, a quantitative estimate of the perceptually salient pitches of the fundamental frequency (F0) contour, is superimposed in white over the F0 contour.

tion, and thus any likely interval code, occurs. The criterion for defining this band was that it was bounded by intervals beyond which the total number of cases per interval (across all emotions) dropped below five. The distribution band of interest was thus between −500 cents and 700 cents.

The sadness distribution was significantly different from a flat distribution, $\chi^2(12) = 67.94$, $p < .01$; as well as from the distribution of intervals in the other three emotions combined, $\chi^2(12) = 119.73$, $p < .01$. What is immediately striking from Figure 1 is the peak at −300 cents for the sadness distribution. This corresponds to a descending minor third. The sadness distribution is also marked by a narrow band of descending intervals (from −500 to −100 cents), with very few intervals outside this band.

The anger distribution is also significantly different from a flat distribution, $\chi^2(12) = 21.09$, $p < .05$; as well as from the distribution of intervals in the other three emotions combined, $\chi^2(12) = 50.15$, $p < .01$. The angry speech samples tend to go up in pitch, and the distribution of intervals is bimodal, with peaks at 100 cents (an ascending minor second) and 700 cents (an ascending perfect fifth).

The pleasantness and happiness distributions are less distinctive. The pleasantness distribution is not significantly different from a flat distribution, $\chi^2(12) = 16.08$, $p > .05$; but is significantly different from the distribution of intervals in the other three emotions combined, $\chi^2(12) = 33.01$, $p < .01$. The happiness distribution shows little discernable structure. It is neither significantly different from a flat distribution, $\chi^2(12) = 5.43$, $p > .05$; nor from the distribution of intervals in the other three emotions combined, $\chi^2(12) = 16.02$, $p > .05$.

Additional acoustic analyses were conducted on the data, examining the following acoustic parameters: speech sample duration, mean intensity, mean F0, the degree of pitch change up or down within a speech sample (with downward pitch change coded as a negative number, measured in cents), and the absolute value (i.e., regardless of direction) of the degree of pitch change within a speech sample (measured in cents). We chose to measure duration, mean intensity, and mean F0 because these are among the most common acoustic parameters measured in previous research (for review, see Juslin & Laukka, 2003). We have used cents to conduct three separate measures of pitch change within a contour. Although these three measures are a function of the same value in cents (i.e., the change in pitch from Syllable 1 to Syllable 2), each is intended to measure a different perceptual characteristic of the pitch contour. We make no assumptions as to whether the pitch

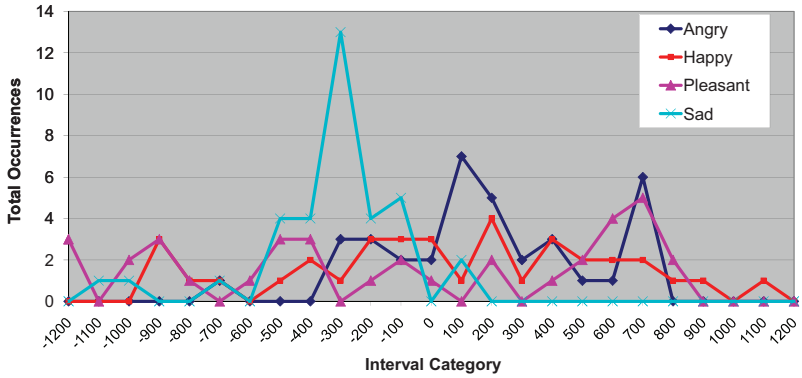

Figure 2. Distribution of speech intervals across interval categories (in cents), grouped according to the intended emotion of the speaker. Negative numbers indicate downward changes in pitch. The mode of the sad distribution is 300 cents downward (minor third), and the angry distribution has bimodal peaks at 100 and 700 cents upward (minor second and perfect fifth).

changes in speech are best measured by a continuum that encodes direction, by a continuum that does not encode direction, or by perceptually based categories. These three measures are intended to reflect three separate levels of auditory perception and representation. The degree of pitch change up or down measures the size of the pitch change between syllables on a continuous scale and takes direction into account by using both negative and positive numbers (for downward and upward pitch changes, respectively). The absolute value of the degree of pitch change is a continuous measure that does not take pitch direction into account but simply captures whether the leap from Pitch 1 to Pitch 2 was small or large. The categorical approach, the results of which are reported above, reflects the categorical framework that tends to dominate our perception of pitch in a musical context.

We conducted separate within-participant analyses of variance (ANOVAs) to determine whether the speech samples differed according to the intended emotion of the speaker for each acoustic parameter. Each participant's parameter means for each of the four emotions (collapsed across the four sentences) were used as the data points for each ANOVA. One speech sample was omitted from analyses involving F0, as the speech sample was whispered and had no F0 contour.

There was a significant main effect of emotion for each of the parameters tested, which were as follows: speech sample duration, $F(3, 24) = 26.23$, $p < .001$; mean intensity, $F(3, 24) = 30.15$, $p < .001$; mean F0, $F(3, 24) = 23.95$, $p < .001$; degree of pitch change up or down within each speech sample (measured in cents), $F(3, 24) = 9.41$, $p < .001$; and the absolute value of the degree of pitch change within each speech sample (in cents), $F(3, 24) = 9.01$, $p < .001$. The parameter means for each emotion are shown in Table 1, along with the results of linearly independent pairwise comparisons (with Bonferroni correction for multiple comparisons). The pairwise comparisons identified the emotions that differed significantly on each acoustic parameter. Angry utterances were significantly longer in duration than pleasant ($p = .002$) and sad ($p < .001$) utterances, and happy utterances were also significantly longer than pleasant ($p = .006$) and sad ($p = .002$) utterances. Angry utterances had a significantly greater mean intensity than pleasant ($p = .018$) and sad utterances ($p = .001$) utterances, and happy utterances also had a significantly greater mean intensity than pleasant ($p = .015$), and sad utterances ($p < .001$). Mean F0 was significantly higher for angry utterances than for sad ones ($p = .018$) and was significantly higher for happy utterances than for angry ($p = .028$), pleasant ($p = .001$), and sad ones ($p < .001$). The degree of pitch change was considered in two ways. First, pitch direction was preserved by coding rising pitch as a positive value and falling pitch as a negative value. The mean size of pitch change was positive for angry and happy utterances and negative for pleasant and sad utterances. Pitch change values were significantly higher for angry utterances than for sad ones ($p < .001$) and significantly higher for pleasant utterances than for sad ones ($p = .037$). In the second method for coding degree of pitch change, the absolute value of pitch change was analyzed. The absolute value of pitch change was significantly greater for pleasant utterances than for angry ($p = .016$) or sad ones ($p = .047$).

Discussion

Pitch intervals seem to be used to encode emotion. This is particularly true of the two negative emotions studied. The distribution of intervals used to express sadness was the most focused of all the emotions, with a peak at −300 cents (a descending minor third). The angry distribution was the next most distinctive, with two peaks: 100 cents (an ascending minor second) and 700 cents (an ascending perfect fifth).

Three findings emerge from this study. First, the negative emotions—sadness and anger—seem to use intervals as communicative devices. One may speculate that negative emotions need more urgently to be detected, because they communicate that something is wrong and may require intervention. It is possible that the costs of failing to detect is higher for negative than for positive emotions.

The second finding is that of the two negative emotions, the low-arousal emotion (sadness) elicited a descending pitch, whereas the high-arousal emotion (anger) elicited an ascending pitch, which is consistent with previous qualitative F0 contour assessments for these emotions (Fónagy, 1978; Juslin & Laukka, 2001; Scherer & Oshinsky, 1977; Williams & Stevens, 1972). It is possible that pitch contour direction may be associated with arousal level, with higher arousal emotions ascending in pitch and lower arousal emotions descending in pitch. However, the positive emotional vocalizations did not show clear patterns of directional specificity, suggesting that there may not be a simple mapping between pitch direction and arousal. We suggest the possibility that directionally specific intervals are used as communicative codes with formal characteristics that are not linearly linked to the physiological effort of expressing negative emotions.

The coding hypothesis is further supported by the third finding, namely, that specific intervals bands are used. This is particularly revealing in the case of anger, for which the distribution is bimodal.

Table 1
Mean Value of Each Acoustic Parameter According to the Emotional State of the Speaker

Acoustic parameter	Angry	Happy	Pleasant	Sad
Duration (in milliseconds)	812	717	528	544
Mean intensity (in decibels)	83	82	76	70
Mean F0 (in hertz)	324	394	283	250
Degree of pitch change (in cents)	182	29	−35	−364
Absolute value of degree of pitch change (in cents)	304	460	749	376

Note. F0 = fundamental frequency.

It is unknown whether listeners use intervals to decode emotion in speech. Experiment 2 was designed to assess the utility of intervals, as well as the other acoustic parameters measured in Experiment 1, as cues by which to decode emotion in speech.

Experiment 2

Participants rated the emotions they perceived in the speech samples from Experiment 1. The primary goal was to assess the acoustic variables that led to emotional identification. Participants listened to each speech sample and rated the degree of perceived anger, happiness, pleasantness, and sadness.

Method

Participants. Ten volunteers (4 male, 6 female) were recruited from the Tufts University community (mean age, 18.7 years). Nine were native speakers of American English who did not grown up in a multilingual household. One spoke Polish until the age of 5, at which point American English became his primary language. Volunteers were compensated for their time with credit for an introductory psychology course.

Stimuli. The 144 speech samples that were collected in Experiment 1 were used as stimuli. The speech samples were presented as stereo digital audio files. To ensure that all stimuli could be presented at a comfortable listening volume, we adjusted the intensity of a few stimuli so that no stimulus was too loud or too soft. (These adjusted intensity values were used in the regressions associated with this experiment.)

Apparatus. Stimuli were presented on a Dell computer with the stimulus presentation software Presentation (Neurobehavioral Systems, Albany, CA). Sound files were played at a comfortable volume over Sony MDR-V600 headphones. The participants made their ratings by using a mouse to click on rating scales that were displayed on a CRT monitor.

Procedure. Each participant was tested individually in a sound-attenuated room. Participants were asked to rate how strongly each speech sample conveyed anger, happiness, pleasantness, and sadness on a scale ranging from 1 (*not at all*) to 7 (*very strongly*). Separate rating scales were used for each emotion, and all four scales were displayed on the computer screen for the duration of each trial. Participants made each rating by using a mouse to click on the appropriate rating box. A speech sample was played automatically at the beginning of each trial. The user interface allowed the participants to replay the speech sample as many times as they desired. The trial ended once a rating had been made on each of the four emotional rating scales.

Results

We used stepwise linear regression to determine which acoustic parameters account for the emotional ratings with the highest degree of consistency. The group mean was calculated for each speech sample on each of the four emotional rating scales. A separate regression was conducted for each rating scale. This analysis takes a categorical approach to conceptualizing the pitch changes in the speech samples. Speech samples were binned into musical interval categories using the perceptual categorizations obtained in Experiment 4. There were 26 speech intervals that could not be significantly categorized in Experiment 4, which were omitted from these analyses. There were six intervals that were larger than one octave, and these were normalized to octave-equivalent categories by reducing interval size by one octave. Normalization through octave equivalence is standard practice in the conceptualization of pitch class hierarchies in tonal music, and we adopted this approach so as to reduce the total number of independent variables in the regression.

Regression determines the utility of the speech intervals as cues for identifying the emotional state of the speaker relative to other acoustic cues. The following acoustic cues were entered as independent variables for each regression: speech sample duration, mean intensity, mean F0, degree of pitch change (up or down) within a speech sample, absolute value of the degree of pitch change, and 25 dummy variables corresponding to interval categories (1,200 cents down, 1,100 cents down, 1,000 cents down, 900 cents down, 800 cents down, 700 cents down, 600 cents down, 500 cents down, 400 cents down, 300 cents down, 200 cents down, 100 cents down, 0 cents, 100 cents up, 200 cents up, 300 cents up, 400 cents up, 500 cents up, 600 cents up, 700 cents up, 800 cents up, 900 cents up, 1,100 cents up, and 1,200 cents up). Table 2 lists the predictors of each regression model in order of their significance and shows the R^2 change associated with the addition of each predictor to the regression model.

Sadness ratings. The mean sadness rating for each stimulus was used as the dependent factor. Seven factors were correlated with the sadness ratings, $R = .612$, $R^2 = .375$, $F(7, 111) = 9.51$, $p < .001$. Significant predictors include 300 cents down (minor third), 100 cents down (minor second), 1,100 cents down (major seventh), mean intensity, absolute degree of pitch change, 500 cents down (perfect fourth), and 700 cents down (perfect fifth).

Table 2
Acoustic Predictors of Variance in the Emotional Ratings of the Speech Samples

Model	Anger	Happiness	Pleasantness	Sadness
1	Duration (+), $R^2 = .319$	Mean F0 (+), $R^2 = .139$	Abs. change (+), $R^2 = .206$	300 Down (+), $R^2 = .134$
2	100 Up (+), $R^2 = .052$	Abs. change (+), $R^2 = .127$	0 Cents (+), $R^2 = .07$	100 Down (+), $R^2 = .07$
3	200 Up (+), $R^2 = .054$	0 Cents (+), $R^2 = .069$	Duration (−), $R^2 = .058$	1,100 Down (+), $R^2 = .039$
4	Pitch change (+), $R^2 = .02$	1,200 Down (+), $R^2 = .029$	600 Up (+), $R^2 = .047$	Mean intensity (−), $R^2 = .032$
5		Duration (−), $R^2 = .034$	Mean F0 (+), $R^2 = .029$	Abs. change (−), $R^2 = .035$
6		900 Down (+), $R^2 = .025$	1,200 Down (+), $R^2 = .027$	500 Down (+), $R^2 = .031$
7		600 Up (+), $R^2 = .022$	400 Down (+), $R^2 = .02$	700 Down (+), $R^2 = .033$

Note. Nature of the correlation, positive (+) or negative (−), is indicated next to the predictor name. Intervals are listed according to their size (in cents) and direction. F0 = fundamental frequency; Abs. change = absolute value of pitch change.

Anger ratings. The mean anger rating for each stimulus was used as the dependent factor. Four factors were correlated with the anger ratings, $R = .667$, $R^2 = .444$, $F(4, 114) = 22.8$, $p < .001$. Significant predictors include speech sample duration, 100 cents up (minor second), 200 cents up (major second), and degree of pitch change.

Happiness ratings. The mean happiness rating for each stimulus was used as the dependent factor. Seven factors were correlated with the happiness ratings, $R = .666$, $R^2 = 0.444$, $F(7, 111) = 12.66$, $p < .001$. Significant predictors include mean F0, the absolute degree of pitch change, 0 cents (unison), 1,200 cents down (octave), duration, 900 cents down (major sixth), and 600 cents up (diminished fifth).

Pleasantness ratings. The mean pleasantness rating for each stimulus was used as the dependent factor. Seven factors were correlated with the pleasantness ratings, $R = .676$, $R^2 = .457$, $F(7, 111) = 13.37$, $p < .001$. Significant predictors include the absolute degree of pitch change, 0 cents (unison), duration, 600 cents up (diminished fifth), mean *F*0, 1,200 cents down (octave), and 400 cents down (major third).

Discussion

The size of the speech intervals clearly accounts for variance in the emotional ratings of sadness, anger, happiness, and pleasantness. For the sadness ratings, the interval of 300 cents down (the minor third) is the most significant predictor, followed by the interval of 100 cents down (the minor second). In music, the interval of 300 cents is associated with sadness, and 100 cents is associated with negative emotions (Cooke, 1959), which is consistent with our findings. Notably, these intervals are more consistent predictors of perceived sadness than all other acoustic factors tested, suggesting that intervals are important components of the acoustic code for the vocal communication of sadness.

Intervals appear to be less crucial in the communication of anger, happiness, and pleasantness, relative to other acoustic factors. The anger ratings were positively associated with 100 cents up (minor second) and 200 cents up (major second), but speech sample duration predicted a much larger proportion of variance in the anger ratings (31.9%, compared with 5.2% and 5.4%, respectively). The happiness and pleasantness ratings were positively associated with the interval categories of 0 cents (unison), 600 cents (diminished fifth), and 1,200 cents (octave). However, other acoustic variables, such as the absolute value of pitch change, were more consistent predictors of perceived happiness and pleasantness. Thus, a pitch change that is relatively large in size, as conceptualized on a continuum, seems to be a better predictor of perceived happiness and pleasantness than intervals, or specific bands of pitch change. Thus, it appears that intervals play a greater role in communicating sadness than other emotions.

Experiment 3 was designed to assess the emotional perception of musical intervals, so as to directly compare the use and perception of intervals across the domains of speech and music. Given that the speech sample pitch contours contained intervals that were generally mistuned, according to the equal-tempered tuning system, the intervals used in Experiment 3 were tuned to match the intervals of each speech sample, so as to allow a direct comparison across domains.

Experiment 3

Melodic musical intervals were presented to participants and rated for perceived emotion. The intervals tested in this experiment were derived from the pitch contour analysis of the speech samples from Experiment 1. Each speech interval was synthesized as a musical interval played with a piano timbre. The musical intervals were matched in size (in cents) to the speech intervals, but all other acoustic parameters were controlled across the musical intervals. We asked participants to rate each interval for the degree of perceived anger, happiness, pleasantness, and sadness, using the same procedure as in Experiment 2.

Method

Participants. Twenty-seven participants were recruited from the Tufts University community and were compensated for their time with credit toward an introductory psychology course. The data from three participants were excluded from analysis because each of these participants deviated from the standard procedure.

The remaining 24 participants comprised 13 females and 11 males. Their mean age was 19.1 years. Sixteen of the participants had played a musical instrument for 1 year or more ($M = 8.3$ years) and had a mean of 6.4 years of formal training on an instrument.

Stimuli. The stimuli were 143 melodic musical intervals, corresponding to the intervals distilled from speech samples. The interval from the pitch contour of each speech sample was synthesized with a musical timbre. Given that the precise interval values from the speech samples generally did not correspond to perfectly tuned musical interval category prototypes, most of these intervals were mistuned, according to the standards of equal-tempered tuning. In addition to these mistuned intervals, 25 interval category exemplars (i.e., perfectly tuned intervals) were used as stimuli. Ascending and descending versions of each musical interval from 100 cents to 1,200 cents were included, as well as the interval of 0 cents, which corresponds to no change in pitch. Each interval was synthesized with a Midi piano timbre using the musical composition software Sibelius 3.0 (Sibelius USA, Inc) at a tempo of 100 beats per minute (i.e., each tone was 600 ms in duration). The intervals were edited with Sony Sound Forge 8.0 (Sony Media Software).

To control for tone height, we kept one tone the same across all of the stimuli. Middle C (262 Hz) was chosen as the common tone. For intervals that went upward in pitch, C was the first tone of the interval. For intervals that went downward in pitch, C was the second tone of the interval. Thus, every interval either began or ended on middle C, and C was the lower tone of every interval. The reason for choosing to keep the lower tone of each interval the same is that it afforded the most effective means of normalizing tone height.

Apparatus. Stimuli were presented on a Dell computer using the stimulus presentation software Presentation (Neurobehavioral Systems). Sound files were played at a comfortable volume over Sony MDR-V600 headphones. The participants made their ratings by using a mouse to click on rating scales that were displayed on a CRT monitor.

Procedure. One musical interval was presented during each trial. Participants were instructed to listen to the interval and gather

their emotional impressions evoked by the interval. They were instructed to rate how strongly each interval conveyed happiness, anger, pleasantness, and sadness on a scale ranging from 1 (*not at all*) to 7 (*very strongly*). They rated each interval on a separate scale for each of the four emotions, and the four scales were shown on the computer screen at the same time. The participants were instructed to indicate multiple emotions in their ratings if the interval reminded them of multiple emotions. The user interface allowed them to replay each interval as many times as they desired during the rating process.

Results

To examine the emotions associated with interval categories, a categorical approach was adopted. The interval categorizations were the same as those used in Experiment 2, based on the results of perceptual testing (Experiment 4). Intervals that could not be significantly categorized in Experiment 4 were omitted from this analysis.

Table 3 lists the mean ratings of perceived anger, happiness, pleasantness, and sadness for each interval category. We conducted a stepwise linear regression on the ratings for each emotional scale to determine which interval categories account most consistently for the variance in the emotional ratings. The group mean ($n = 24$) was calculated for each interval (i.e., for each individual stimulus) on each of the four emotional rating scales. The interval stimuli were treated categorically and coded as dummy variables. The following interval categories were used as dependent factors in each regression: 1,200 cents down, 1,100 cents down, 1,000 cents down, 900 cents down, 800 cents down, 700 cents down, 600 cents down, 500 cents down, 400 cents down, 300 cents down, 200 cents down, 100 cents down, 0 cents, 100 cents up, 200 cents up, 300 cents up, 400 cents up, 500 cents up, 600 cents up, 700 cents up, 800 cents up, 900 cents up, 1,100 cents up, and 1,200 cents up.

The regressions identify the interval categories that are most consistently associated with the high and low scores on each emotional scale. Table 4 lists the intervals that predicted variance in the emotional ratings. The predictors for each regression model are listed in order of their significance, and the R^2 change associated with each predictor is shown.

Sadness ratings. The mean sadness rating for each stimulus was used as the dependent factor. Fifteen intervals were correlated with the sadness ratings, $R = .914$, $R^2 = .835$, $F(15, 128) = 43.2$, $p < .001$; including the following intervals, which were positively correlated with the sadness ratings and accounted for sizable proportions of variance: 100 cents down (24.7%), 300 cents down (20.2%), and 100 cents up (7.3%).

Anger ratings. The mean anger rating for each stimulus was used as the dependent factor. Thirteen intervals were correlated with the anger ratings, $R = .936$, $R^2 = .876$, $F(13, 130) = 70.36$, $p < .001$; including the following intervals, which were positively correlated with the anger ratings and accounted for sizable proportions of variance: 100 cents up (29.1%), 100 cents down (22.1%), 600 cents up (13.2%), and 300 cents down (6.7%).

Happiness ratings. The mean happiness rating for each stimulus was used as the dependent factor. Seventeen intervals were correlated with the happiness ratings, $R = .945$, $R^2 = .892$, $F(17, 126) = 61.35$, $p < .001$; including the following intervals, which were positively correlated with the happiness ratings and accounted for sizable proportions of variance: 700 cents up (27.7%) and 900 cents up (5.7%).

Table 3
Mean Ratings of Perceived Emotion for Each Interval Category

Interval	Anger	Happiness	Pleasantness	Sadness
0 cents (Unison)	2.96	2.31	2.72	3.66
100 cents	4.26 (D)	1.68 (D)	2.14 (D)	4.82 (D)
(Minor 2nd)	4.72 (U)	1.81 (U)	2.35 (U)	4.19 (U)
200 cents	2.77 (D)	2.47 (D)	2.98 (D)	3.79 (D)
(Major 2nd)	2.70 (U)	2.75 (U)	3.47 (U)	3.32 (U)
300 cents	3.39 (D)	2.09 (D)	2.61 (D)	4.47 (D)
(Minor 3rd)	3.08 (U)	2.22 (U)	2.99 (U)	4.10 (U)
400 cents	2.58 (D)	3.09 (D)	3.52 (D)	3.28 (D)
(Major 3rd)	2.31 (U)	3.29 (U)	3.82 (U)	3.19 (U)
500 cents	2.42 (D)	3.35 (D)	3.92 (D)	3.20 (D)
(Perfect 4th)	2.41 (U)	3.93 (U)	3.63 (U)	2.78 (U)
600 cents	3.96 (D)	1.98 (D)	2.21 (D)	4.44 (D)
(Diminished 5th)	4.34 (U)	2.49 (U)	2.42 (U)	3.56 (U)
700 cents	1.82 (D)	4.07 (D)	4.74 (D)	3.05 (D)
(Perfect 5th)	2.45 (U)	4.47 (U)	4.28 (U)	2.45 (U)
800 cents	2.85 (D)	2.63 (D)	3.08 (D)	4.07 (D)
(Minor 6th)	3.34 (U)	3.12 (U)	3.04 (U)	3.75 (U)
900 cents	2.63 (D)	3.39 (D)	3.54 (D)	3.31 (D)
(Major 6th)	2.00 (U)	4.92 (U)	4.46 (U)	2.65 (U)
1,000 cents	2.80 (D)	2.92 (D)	3.02 (D)	3.59 (D)
(Minor 7th)	2.75 (U)	3.63 (U)	4.13 (U)	3.50 (U)
1,100 cents	3.61 (D)	2.29 (D)	2.03 (D)	3.94 (D)
(Major 7th)	4.21 (U)	2.57 (U)	2.50 (U)	3.55 (U)
1,200 cents	2.23 (D)	4.03 (D)	3.94 (D)	2.93 (D)
(Octave)	2.42 (U)	4.96 (U)	4.38 (U)	2.29 (U)

Note. D = downward intervals; U = upward intervals.

Pleasantness ratings. The mean pleasantness rating for each stimulus was used as the dependent factor. Fifteen intervals were correlated with the pleasantness ratings, $R = .866$, $R^2 = .75$, $F(15, 128) = 25.54$, $p < .001$; including the following intervals, which were positively correlated with the pleasantness ratings and accounted for sizable proportions of variance: 700 cents up (17%) and 700 cents down (6.2%).

Discussion

Although each regression identified numerous intervals that predicted the emotional ratings, only a few interval categories emerged as consistent predictors of a large proportion of variance in the emotional ratings. Notably, the interval of 100 cents (the minor second, both ascending and descending) was positively linked to perceived anger and perceived sadness and was negatively correlated with perceived happiness and pleasantness. The interval of 300 cents downward (the minor third) was also strongly linked to perceived sadness and features as the fourth most significant predictor for anger (positively correlated), happiness (negatively correlated), and pleasantness (negatively correlated). The interval of 600 cents upward (the diminished fifth) was also positively correlated with anger. The interval of 700 cents upward (the perfect fifth) was positively correlated with happiness and

Table 4
Predictors of Emotion Ratings of Musical Intervals

Model	Anger	Happiness	Pleasantness	Sadness
1	100 Up (+), R^2 = .291	700 Up (+), R^2 = .277	700 Up (+), R^2 = .17	100 Down (+), R^2 = .247
2	100 Down (+), R^2 = .221	100 Down (−), R^2 = .119	100 Down (−), R^2 = .12	300 Down (+), R^2 = .202
3	600 Up (+), R^2 = .132	100 Up (−), R^2 = .095	100 Up (−), R^2 = .08	700 Up (−), R^2 = .15
4	300 Down (+), R^2 = .067	300 Down (−), R^2 = .099	300 Down (−), R^2 = .074	100 Up (+), R^2 = .073
5	1,100 Up (+), R^2 = .046	900 Up (+), R^2 = .057	700 Down (+), R^2 = .062	500 Up (−), R^2 = .027
6	600 Down (+), R^2 = .034	500 Up (+), R^2 = .036	600 Up (−), R^2 = .05	600 Down (+), R^2 = .024
7	800 Up (+), R^2 = .021	1,200 Down (+), R^2 = .039	1,100 Down (−), R^2 = .037	300 Up (+), R^2 = .017
8	1,100 Down (+), R^2 = .02	1,200 Up (+), R^2 = .034	0 Cents (−), R^2 = .03	800 Down (+), R^2 = .017
9	700 Down (−), R^2 = .017	700 Down (+), R^2 = .036	600 Down (−), R^2 = .032	200 Down (+), R^2 = .018
10	0 Cents (+), R^2 = .008	500 Down (+), R^2 = .019	200 Down (−), R^2 = .023	1,200 Up (−), R^2 = .014
11	300 Up (+), R^2 = .008	900 Down (+), R^2 = .021	1,100 Up (−), R^2 = .023	900 Up (−), R^2 = .013
12	900 Up (−), R^2 = .006	400 Down (+), R^2 = .013	900 Up (+), R^2 = .017	1,200 Down (−), R^2 = .011
13	1,200 Down (−), R^2 = .004	400 Up (+), R^2 = .015	500 Down (+), R^2 = .014	1,100 Down (+), R^2 = .007
14		800 Up (+), R^2 = .01	1,200 Down (+), R^2 = .009	800 Up (+), R^2 = .007
15		1,000 Up (+), R^2 = .01	1,200 Up (+), R^2 = .009	0 Cents (+), R^2 = .007
16		1,000 Down (+), R^2 = .006		
17		200 Up (+), R^2 = .007		

Note. The nature of the correlation, positive (+) or negative (−), is indicated next to the predictor name. Intervals are listed according to their size (in cents) and direction.

pleasantness, accounting for the largest proportion of variance for each of those emotions.

The emotional ratings of 100, 300, 600, and 700 cents were generally consistent with the historic associations and the reported empirical assessments of melodic intervals. The interval of 100 cents (the minor second) has been described as denoting "spiritless anguish" (Cooke, 1959, p. 89) and is perceived as melancholy and tense (Maher & Berlyne, 1982), which is consistent with this interval's strong mappings to anger and sadness evinced in the present experiment. The interval of 300 cents (the minor third) denotes "tragedy" (Cooke, 1959, p. 89) and is perceived as melancholy (Maher & Berlyne, 1982), which is consistent with the current findings. The interval of 600 cents (the diminished fifth) denotes "devilish and inimical forces" (Cooke, 1959, p. 89) and is perceived as tense, displeasing, and ugly (Maher & Berlyne, 1982); these descriptive assessments lack the emotional specificity of the evinced mapping between this interval and anger, but all of these assessments correspond to negative affective states with high levels of arousal. Finally, the interval of 700 cents (the perfect fifth) is perceived as pleasing and beautiful (Maher & Berlyne, 1982), which is consistent with the evinced mapping between this interval and positive valence.

The present findings suggest that the mappings between intervals and emotions extend beyond interval category prototypes (i.e., perfectly tuned intervals). The intervals in the present experiment were generally mistuned and thus could not be considered ideal interval category exemplars. Nonetheless, the emotional assessments of these mistuned intervals were consistent with the previously reported emotional associations with interval category exemplars. These findings suggest that imprecise tuning does not change the overall emotion conveyed by an interval relative to the emotion conveyed by its category exemplar. These findings have obvious implications for the emotional perception of pitch contours in speech. The pitch contours analyzed in Experiment 1 revealed that the intervals in speech are generally mistuned from musical interval category exemplars. However, the results of the present experiment suggest that emotional associations are consistent within an interval category, regardless of mistuning. Thus, the imperfect tuning of intervals in speech is not problematic for the categorical conceptualization and assessment of these intervals, as the emotional associations that exist for musical intervals extend to mistuned intervals that fall within the perceptual bounds of an interval category.

The previous experiments utilized a categorical approach to the assessment of two-tone pitch sequences in speech and music. This categorical approach is supported by perceptual data on interval perception (Burns & Campbell, 1994; Burns & Ward, 1978; Siegel & Siegel, 1977a, 1977b) and was necessitated by the goal of this line of investigations, which was to determine whether the pitch patterns used in speech to communicate emotion were similar to the pitch patterns used in music to communicate the same emotions. Previous research suggests that the perceptual category boundary occurs at 50 cents between interval category prototypes. Our stimuli contained intervals that closely bordered the perceptual category boundary, so it seemed necessary to obtain perceptual categorizations of the intervals to determine the consistency with which an interval is perceived as belonging to particular category across individuals. Given that the regression analyses in Experiments 2 and 3 included interval categorizations, Experiment 4 was conducted to obtain perceptual verification of these interval categorizations. The results of this experiment allowed us to identify intervals that were perceptually ambiguous, because of close proximity to the perceptual category boundary, and to omit those intervals from the Experiment 2 and 3 regressions.

Experiment 4A

This investigation was designed to test the perceptual categorization of musical intervals distilled from the speech samples. The results of this experiment were used to bin speech samples and musical intervals into interval categories, as necessitated by the regression analyses in Experiments 2 and 3. Additionally, this

investigation allowed us to identify intervals that could not be categorized consistently across individuals and exclude those intervals from our Experiment 2 and 3 regressions, as a categorical approach is unjustifiable for perceptually ambiguous intervals. Thus, we consider this to be a supporting experiment, as it was conducted solely to obtain the perceptual categorizations used in the Experiment 2 and 3 regressions.

We use the term *categorization* to refer to the explicit assignment of a mistuned interval to a musical interval category, as determined by a forced-choice perceptual similarity judgment in which the mistuned interval is compared to interval category exemplars from the two nearest interval categories. Although interval categorization is likely to occur implicitly when one encounters mistuned musical intervals under naturalistic conditions of music perception, our use of the term in the description of this experiment refers to the explicit assessment of an interval. Previous research suggests that musical intervals are perceived categorically (Burns & Campbell, 1994; Burns & Ward, 1978; Siegel & Siegel, 1977a, 1977b), with the category boundary between semitones occurring at 50 cents. Experiments that have tested the categorization of musical intervals have chosen tasks that may have led to categorizations that were based largely on the perceived size of the interval. However, it is possible that the emotional quality of an interval may be a feature that can be used for categorization, and it is possible that the category boundaries may differ if one is asked to categorize intervals based on their emotional similarity, rather than their size. Thus, the participants in this investigation were asked to base their similarity judgments on the emotional quality of the intervals. Each interval was synthesized with a piano timbre. On each trial, the participant was asked to compare a test interval (distilled from a speech sample and most often mistuned from the musical interval category prototype) to two precise musical intervals that corresponded to prototypes from the two interval categories nearest to the test interval. The participant was then asked to decide which of the comparison intervals was emotionally more similar to the test interval.

Method

Participants. Ten volunteers were recruited from the Tufts University community (4 female, 6 male). Their mean age was 19.6 years. Six of the participants identified themselves as musicians, with an average of 4 years of formal training on their primary instrument (range = 0–9 years) and an average of 6.6 years of experience playing their primary instrument (range = 5–10). Volunteers were compensated for their time with credit for an introductory psychology course.

Stimuli. A total of 143 test intervals and 39 comparison intervals were synthesized for this experiment, which entailed 143 trials. The speech samples obtained in Experiment 1 were the basis of the test stimuli. The test intervals were the same stimuli as those used in Experiment 3. The comparison intervals were synthesized to match the test intervals in timbre and tone duration, using the same software described in Experiment 3. Each comparison interval corresponded to a precise equal-tempered musical interval prototype (i.e., 100 cents, 200 cents, 300 cents, etc.). Because some test intervals went up in pitch and others went down, the pitch direction of the comparison intervals for each trial was matched to the pitch direction of the test interval. All intervals were normalized to the same tone (the lowest tone in every interval was a middle C).

Apparatus. Stimuli were presented on a Dell computer using the stimulus presentation software Presentation (Neurobehavioral Systems). Sound files were played at a comfortable volume over Sony MDR-V600 headphones. The participants made their ratings by using keys on a computer keyboard.

Procedure. Each participant was seated in a sound-attenuated room. They were told that on each trial they would hear four pairs of tones and would be prompted to decide which pairs were more similar: Pairs 1 and 2 or Pairs 3 and 4. Each trial began with the test interval. While the test interval was played over headphones, the word "Test" was displayed on a CRT monitor. The pair of test tones was followed by a comparison set of tones. The words "Sound 1" were displayed on the monitor while this pair of comparison tones was played. Then, the test interval was played again, accompanied by the word "Test" on the monitor. This was followed by a second set of comparison tones, which were accompanied by the words "Sound 2" on the monitor. Participants then decided whether Sound 1 or Sound 2 was more similar to the test tones, by pressing a key on a computer keyboard. The participants were instructed that they should listen to the emotional quality of the test interval and comparison intervals and that they should use their impressions of the overall emotional similarity between the test and comparison intervals to make their judgments.

Results

A binomial distribution was the basis for determining whether the level of interparticipant agreement for the categorization of each stimulus was significantly greater than what would occur by chance. An individual participant's response to each stimulus was considered an independent observation, resulting in 10 independent observations for each stimulus (one observation for each participant). For each stimulus, the chance probability of choosing either response was 50%. Categorizations were deemed significant (at $p < .044$) if at least 8 of 10 participants agreed on the categorization of an individual stimulus.

The participants were able to significantly categorize 88.1% of the stimuli. All of the participants agreed about the categorization of 58% of the stimuli. Only 11.9% of the stimuli were not categorized significantly above chance.

The following intervals could not be significantly categorized: 154, 155, 157 down; 164, 165, 167, 250, 346, 349, 356, 364 up; 438, 458, 551, 635, 655, and 855 cents. The direction of an interval is indicated for intervals that categorized significantly in one direction but failed to reach significance in the other direction (e.g., 157 and 364 cents).

Discussion

The significant interval categorizations are consistent with the categorization boundaries reported in the literature on categorization of intervals, with the exception of one interval. The interval of 44 cents was significantly categorized as 100 cents, rather than as 0 cents. The predicted category boundary is 50 cents, on the basis of previous findings. Because the categorization of this interval is inconsistent with the predicted categorization, this interval was included in a follow-up investigation.

There were 17 intervals that could not be categorized significantly above chance. A follow-up experiment was designed to focus on the categorization of these 17 intervals in particular, as well as the aforementioned interval.

Experiment 4B

This investigation was designed to further test the categorization of musical intervals. Experiment 4A resulted in the categorization of 88.11% of the intervals tested. This investigation focuses on obtaining categorizations for the intervals that were not significantly categorized, as well as obtaining additional categorization trials for the intervals that were categorized on the borderline of significance. Each of the uncategorized intervals from Experiment 4A was used as a stimulus in this experiment, and each was presented randomly three times in the experiment.

Method

Participants. Five volunteers were recruited from the Tufts University community (all 5 were male). Their mean age was 18.6 years. Four of the participants identified themselves as musicians, with an average of 4.5 years of formal training on their primary instrument (range = 3–6 years) and an average of 7.3 years of experience playing their primary instrument (range = 4–12). The volunteers were compensated for their time with credit for an introductory psychology course.

Stimuli. A subset of the stimuli used in Experiment 4A was used in this experiment. This subset included the 17 intervals that were not significantly categorized in the previous experiment, as well as 13 additional intervals that were in close cent value to those that were not significantly categorized (these additional intervals, though significantly categorized in the previous experiment, generally had p values near the borderline of significance).

Apparatus. Stimuli were presented on a Dell computer using the stimulus presentation software Presentation (Neurobehavioral Systems). Sound files were played at a comfortable volume over Sony MDR-V600 headphones. The participants made their ratings by using keys on a computer keyboard.

Procedure. The trial parameters were the same as in the previous experiment, as were the procedure and task. However, each stimulus in this experiment was presented a total of three times at random over the course of the experiment (for a total of 90 trials). This repetition should yield a better assessment of how each participant perceived each interval.

Results

The total number of trials for each interval was 15 (three per each of 5 participants). Each trial was treated as an independent observation, with a 50% probability of choosing either response. If the categorization for a particular interval was the same for 11 of 15 trials, the categorization was deemed to be significant (at $p < .043$), as determined by a binomial distribution.

The participants were able to categorize 15 of the 30 intervals at the requisite 11-of-15 trial agreement. The following intervals were not significantly categorized: 44, 155, 157 up; 250, 346, 349, 352, 361, 458, 459, 551, 655, 660, 661, and 855 cents.

Discussion

All of the intervals that were significantly categorized in this experiment were categorized consistently with the category boundaries established by previous research. The interval of 44 cents, which received a categorization in the previous experiment that was inconsistent with the hypothesized categorization, was not categorized at a level of significance in this follow-up investigation, suggesting that it may be inappropriate to reject the hypothesized categorization.

The results of this analysis suggest that the boundary between interval categories is 50 cents, as suggested by previous research. Experiment 4A showed a high degree of interparticipant agreement on the categorization of intervals, particularly those that fell well within the interval category (i.e., ± 35 cents of the interval category prototype). Intervals that fell within a few cents of a category boundary were not always significantly categorized, as indicated by the results of Experiments 4A and 4B. The following intervals failed to reach significance in at least one of the two experiments: 44, 154, 155, 157 down; 157 up; 164, 165 down; 167, 250, 346, 349, 352, 356, 361, 364 up; 438, 458, 459, 551, 635 up; 655, 660, 661, and 855 cents. Failure to significantly categorize these stimuli into interval categories resulted in the exclusion of the speech samples corresponding to these particular intervals from the regression analyses utilized in Experiment 2, as well as the exclusion of these musical intervals from the analyses in Experiment 3. Given that the approach utilized in the analyses of Experiments 2 and 3 binned each interval into a perceptual category, it was necessary to establish the perceptual consistency of those categorical designations. Intervals that could not be consistently categorized at the level of significance in Experiment 4 were omitted from the previously reported regression analyses.

General Discussion

The preceding experiments suggest that human vocal expressions of sadness and anger use pitch patterns that approximate those used in music to convey the same emotions. The pitch patterns that were used similarly across domains to encode sadness were the descending minor third (300 cents) and the descending minor second (100 cents). The ascending minor second (100 cents) was used to encode anger across domains.

Disparities also emerged in the comparison across domains. The ascending interval of 600 cents (the diminished fifth) was positively associated with a small proportion of variance in the happiness and pleasantness ratings of the speech samples. This interval accounted for a significant proportion of variance in the emotional ratings of the musical intervals, but its affective association mapped to the other end of the valence spectrum, as it was perceived to convey anger. The interval of 700 cents (the perfect fifth) is close in size to 600 cents and was strongly associated with positive valence in the musical interval ratings, but it was not significantly associated with valence in the speech ratings. The disparity across domains concerning 600 and 700 cents raises two obvious possibilities. One is that the instances of 600 cents in the happy and pleasant speech samples were failed attempts at producing 700-cent intervals, but successful attempts to communicate the intended emotions, because of the redundancy of multiple acoustic cues mapping to the intended emotion. Another possibil-

ity is that a cross-domain mapping does not exist for intervals in the 600–700-cent range. The small amount of variance accounted for in the speech ratings by the interval of 600 cents points to the greater utility of other acoustic cues in decoding pleasantness and happiness.

It is unknown to what extent these findings generalize to longer utterances or if they occur cross-culturally. However, these findings suggest that the acoustic code for communicating emotion across the domains of music and speech is more similar than previously thought. The clearest correspondence across domains occurred in the communication of sadness. The minor third (300 cents) was the most common interval used to encode sadness in speech. It was also the acoustic feature that predicted the perception of sadness more consistently than any other in the decoding phase of emotional communication. The minor third was also strongly associated with perceived sadness when it was presented as a musical stimulus. Taken together, these results suggest that the minor third is used to communicate sadness in both speech and music. The minor second (100 cents) was used in the communication of negative emotions in both domains and showed a tendency toward directional specificity, with the ascending minor second mapped more strongly to anger than sadness, and the converse pattern shown for the descending minor second.

The present findings may have implications for mental health practitioners. It is possible that the descending minor third and minor second may not only signify sadness but may also be measures of depression. The detection of these intervals in speech has the potential to serve as an objective means of assessing one's emotional state in the absence of explicit self-reports or overt physical cues. The use of these cues by practitioners could result in more comprehensive monitoring of mental health. Further research is needed to assess this possibility.

The discovery of emotion-specific pitch patterns within the F0 contour of speech was obtained with a relatively novel methodological approach to F0 contour analysis. Assessment of the F0 contour of speech is inherently challenging, as there is often tremendous F0 variability, even within short utterances. Steady-state pitches (i.e., portions of the contour in which F0 variability is minimal) are often very short in duration, and the F0 contour tends to drift substantially during the articulatory adjustments necessary for phonemic transitions. It is a challenge to determine which properties of the F0 contour may be of acoustic importance in the process of communication. We used the prosogram model (Mertens, 2004) to quantify the F0 contour data, as it utilizes various acoustic cues (e.g., phonemic boundaries, the intensity contour, and frequency variability thresholds) to model human perception of the F0 contour. That is, it uses perceptually relevant acoustic cues to identify the portions of the F0 contour that are likely to be perceptually salient. We assessed the pitch patterns of the vocal contours using the semitone scale, a musical system of measurement that is more appropriate than using Hz to describe pitch patterns. Although Hz is appropriate for measuring the frequency of a pitch or capturing the mean F0 of a pitch contour, using Hz measurement alone to characterize a pitch pattern can actually mask the perceptual characteristics of the pitch pattern, because pitch perception functions logarithmically in relation to Hz measurement. To illustrate, imagine two people singing "Happy Birthday" independently, each starting on a different pitch. If one were to transcribe the pitches of each vocalization in Hz, it would be very difficult to determine from the transcriptions alone that the singers were actually singing identical pitch patterns. Not only would the patterns contain different frequencies, but the size of the frequency difference between adjacent pitches would also differ across the singers. Only a measurement system that reflects pitch perception in a linear manner can preserve the relationship between pitches and capture the perceptual similarity of these vocalizations. The semitone system of measurement does just that and is the standard system for measuring pitch in music. Thus, using the semitone system of measurement to capture the perceptual relationships within a prosodic pitch contour, in conjunction with Hz measurement, can yield a more comprehensive assessment of pitch contours and elucidate the perceptual similarities between contours. Additionally, the use of this musical system of measurement facilitates comparisons across the domains of speech and music. As the present research has revealed, pitch patterns are used similarly across domains to communicate sadness. It is likely that there are numerous correspondences across the domains of music and speech yet to be discovered.

Evolutionary Origins

The cross-domain correspondences between these pitch patterns and emotional states naturally lead to questions about the evolutionary origins of these associations. We may wonder whether these associations were arbitrary in origin and developed into a formal communicative code that has become widespread or whether these associations reflect automatic affective responses to sounds that arise not from arbitrary associations but from how certain classes of sounds physically interact with the auditory system.

One may experience a visceral affective response to a particular sound as a function of the human auditory system's limited capacity to resolve the spectral properties of the sound. When one is presented with two tones that occur within the same critical band of frequencies, the tones excite similar portions of the basilar membrane and cause physical interference, giving rise to an unpleasant perceptual beating sensation, which is often described as rough and dissonant (e.g., Plomp & Levelt, 1965). Musical dissonance is generally thought to reflect sensory dissonance (although, strictly speaking, the two concepts are not the same; see Bregman, 1990) and is similarly associated with negative affect. Although many listeners find musical dissonance aesthetically appealing, this appreciation may be learned, as it is not evident in infants. A preference for consonant intervals over dissonant intervals has been observed in 2-month-old infants (Trainor, Tsang, & Cheung, 2002) and is well documented in older infants (Trainor & Heinmiller, 1998; Zentner & Kagan, 1996, 1998).

The minor second (100 cents) is considered a dissonant musical interval. Thus, the association between the minor second and negative affect has traditionally been explained as a response to sensory dissonance. Although the physical interference underlying sensory dissonance is thought to occur only for simultaneous tones (referred to musically as *harmonic intervals*), the psychoacoustic association with negativity extends to melodic intervals as well (Maher & Berlyne, 1982). Although the dissonance associated with the harmonic minor second is an obvious mapping source for the association between the melodic minor second and negative valence, the present findings suggest that this association is also

shared with affective speech. The consistent directionality effects for this interval across domains (ascending minor seconds for anger and descending minor seconds for sadness) suggest that the emotional mappings for the melodic minor second extend beyond what can be mapped solely from sensory dissonance, as these directionality effects obviously cannot be mapped from tones that occur simultaneously. Thus, the association between the minor second and negativity may reflect an automatic negative response to dissonance, which is apparent very early in life (Zentner & Kagan, 1996, 1998), coupled with an environmental influence that has specified directionality so that the ascending minor second is associated with anger (and perhaps with other negative high-arousal emotions) and the descending minor second is associated with sadness (and perhaps with other negative low-arousal emotions).

The association between the minor third (300 cents) and sadness is not so easily explained. Unlike the minor second, the minor third is a consonant musical interval. Therefore, its association with sadness cannot be attributed to underlying sensory dissonance. It is likely that this interval has been adopted as a formal code for communicating sadness. This code may not be arbitrary in origin, as it may reflect a vocal response pattern that is shaped by the prototypical alignment of physiological variables underlying the phenomenological experience of sadness. Physiological variables influence aspects of vocal production including phonation, vocal resonance, and subglottal pressure (Johnstone & Scherer, 2000; Scherer, 1989). The particular alignments of physiological variables underlying the range of experiential states identified collectively as "sadness" may shape vocal outputs in such a way that the descending minor third has a high probability of occurrence. It is likely that any pattern that typifies the expression of an emotional state will be preferred—both phylogenetically and ontogenetically—over patterns that are less typical, as selection for a typical pattern will lead to a higher probability of communicative success. Given that the present findings are specific to patterns produced by speakers of American English, it is necessary to examine the prosodic patterns produced across cultures to determine whether the minor third is used universally to communicate sadness. Such findings will elucidate the whether the minor third is a vocal pattern tied to the physiological manifestations of sadness.

The present findings may appear to support the theory that language has served evolutionarily as a mapping source for the musical use of acoustic parameters in the communication of emotion. Although we neither endorse nor attempt to discredit this theory, we want to draw attention to an alternative theory that we regard as highly plausible. It has been proposed that a communicative system comprising acoustic elements common to both language and music was an evolutionary predecessor to language and music and that the two domains evolved divergently from this common origin (e.g., Brown, 2000; Cross & Woodruff, 2009). Dubbed "musilanguage" by Steven Brown (2000), this precursor is theorized to have been characterized by features common to both domains, such as phrases formed from a finite set of pitches and rhythmic elements used in infinite generative combinations. This system is thought to have subserved rudimentary referential and emotive functions. The present findings add to the growing list of structural and functional features common to language and music. Such commonalities suggest the importance of further research and may point to answers regarding the evolutionary origins of emotional communication.

References

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

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International, 5,* 341–345. Available online at www.praat.org

Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound.* Cambridge, MA: The MIT Press.

Brown, S. (2000). The "musilanguage" model of musical evolution. In N. L. Wallin, B. Merker, & S. Brown (Eds.), *The origins of music* (pp. 271–300). Cambridge, MA: The MIT Press.

Burns, E. M., & Campbell, S. L. (1994). Frequency and frequency-ratio resolution by possessors of absolute and relative pitch: Examples of categorical perception? *Journal of the Acoustical Society of America, 96,* 2704–2719.

Burns, E. M., & Ward, W. D. (1978). Categorical perception—Phenomenon or epiphenomenon: Evidence from experiments in the perception of musical intervals. *Journal of the Acoustical Society of America, 63,* 456–468.

Cooke, D. (1959). *The language of music.* London: Oxford University Press.

Cross, I., & Woodruff, G. E. (2009). Music as a communicative medium. In R. Botha & C. Knight (Eds.), *The prehistory of language* (Vol. 1, pp. 113–144). Oxford, England: Oxford University Press.

Crowder, R. G. (1985). Perception of the major/minor distinction: III. Hedonic, musical, and affective discriminations. *Bulletin of the Psychonomic Society, 23,* 314–316.

D'Allesandro, C., & Castellengo, M. (1994). The pitch of short-duration vibrato tones. *Journal of the Acoustical Society of America, 95,* 1617–1630.

Ekman, P., Levenson, R. W., & Friesen, W. V. (1983). Autonomic nervous system activity distinguishes among emotions. *Science, 221,* 1208–1210.

Fónagy, I. (1978). A new method of investigating the perception of prosodic features. *Language and Speech, 21,* 34–49.

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

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

Hevner, K. (1935b). Expression in music: A discussion of experimental studies and theories. *Psychological Review, 47,* 186–204.

Hevner, K. (1936). Experimental studies of the elements of expression in music. *American Journal of Psychology, 48,* 246–268.

Hevner, K. (1937). The affective value of pitch and tempo in music. *American Journal of Psychology, 49,* 621–630.

House, D. (1990). *Tonal perception in speech.* Lund, Sweden: Lund University Press.

Johnstone, T., & Scherer, K. R. (2000). Vocal communication of emotion. In M. Lewis & J. M. Haviland-Jones (Eds.), *Handbook of emotions* (2nd ed., pp. 220–235). New York: Guilford Press.

Jowett, B. (1885). *The politics of Aristotle.* Oxford, England: Oxford at the Clarendon Press.

Juslin, P. N. (1997). Emotional communication in music performance: A functionalist perspective and some data. *Music Perception, 14,* 383–418.

Juslin, P. N. (2001). Communicating emotion in music performance: A review and a theoretical framework. In P. N. Juslin & J. A. Sloboda (Eds.), *Music and emotion: Theory and research* (pp. 309–337). New York: Oxford University Press.

Juslin, P. N., & Laukka, P. (2001). Impact of intended emotion intensity on

decoding accuracy and cue utilization in vocal expression of emotion. *Emotion, 1,* 381–412.

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

Kivy, P. (1989). *Sound sentiment.* Philadelphia: Temple University Press.

Krumhansl, C. L. (1997). An exploratory study of musical emotions and psychophysiology. *Canadian Journal of Experimental Psychology, 51,* 336–352.

Maher, T. F., & Berlyne, D. E. (1982). Verbal and exploratory responses to melodic musical intervals. *Psychology of Music, 10,* 11–27.

Mertens, P. (2004, March 23–26). *The prosogram: Semi-automatic transcription of prosody based on a tonal perception model.* Paper presented at Speech Prosody 2004: International Conference, Nara, Japan. Available at http://www.isca-speech.org/archive/sp2004/sp04_549.pdf

Nielzén, S., & Cesarec, Z. (1982). Emotional experience of music as a function of musical structure. *Psychology of Music, 10,* 7–17.

Papoušek, M. (1996). Intuitive parenting: A hidden source of musical stimulation in infancy. In I. Deliège & J. A. Sloboda (Eds.), *Musical beginnings* (pp. 88–112). Oxford, England: Oxford University Press.

Patel, A. D. (2005). The relationship of music to the melody of speech and to syntactic processing disorders in aphasia. *Annals of the New York Academy of Sciences, 1060,* 59–70.

Patel, A. D., Iverson, J. R., & Rosenberg, J. D. (2006). Comparing the rhythm and melody of speech and music: The case of British English and French. *Journal of the Acoustical Society of America, 119,* 3034–3047.

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

Plomp, R., & Levelt, W. J. M. (1965). Tonal consonance and critical bandwidth. *Journal of the Acoustical Society of America, 38,* 548–560.

Scherer, K. R. (1989). Vocal correlates of emotional arousal and affective disturbance. In H. Wagner & A. Manstead (Eds.), *Handbook of social psychophysiology* (pp. 165–197). New York: Wiley.

Scherer, K. R. (1995). Expression of emotion in voice and music. *Journal of Voice, 9,* 235–248.

Scherer, K. R., & Oshinsky, J. S. (1977). Cue utilisation in emotion attribution from auditory stimuli. *Motivation and Emotion, 1,* 331–346.

Siegel, J. A., & Siegel, W. (1977a). Absolute identification of notes and intervals by musicians. *Perception & Psychophysics, 21,* 143–152.

Siegel, J. A., & Siegel, W. (1977b). Categorical perception of tonal intervals: Musicians can't tell sharp from flat. *Perception & Psychophysics, 21,* 399–407.

Trainor, L. J., & Heinmiller, B. M. (1998). The development of evaluative responses to music: Infants prefer to listen to consonance over dissonance. *Infant Behavior and Development, 21,* 77–88.

Trainor, L. J., Tsang, C. D., & Cheung, V. H. W. (2002). Preference for sensory consonance in 2- and 4-month-old infants. *Music Perception, 20,* 187–194.

Wedin, L. (1972). Multidimensional study of perceptual-emotional qualities in music. *Scandinavian Journal of Psychology, 13,* 241–257.

Williams, C. E., & Stevens, K. N. (1972). Emotions and speech: Some acoustical correlates. *Journal of the Acoustical Society of America, 52,* 1238–1250.

Zentner, M. R., & Kagan, J. (1996). Perception of music by infants. *Nature, 383,* 29.

Zentner, M. R., & Kagan, J. (1998). Infants' perception of consonance and dissonance in music. *Infant Behavior and Development, 21,* 483–492.

Received October 27, 2008
Revision received October 5, 2009
Accepted October 5, 2009 ■