
Polyphonic Voice Multiplicity, Numerosity, and Musical Emotion Perception
Author(s): Yuri BrozeBrandon T. PaulErin T. AllenKathleen M. Guarna
Source: *Music Perception: An Interdisciplinary Journal*, Vol. 32, No. 2 (December 2014), pp. 143-159
Published by: University of California Press
Stable URL: http://www.jstor.org/stable/10.1525/mp.2014.32.2.143
Accessed: 18/12/2014 10:21

POLYPHONIC VOICE MULTIPLICITY, NUMEROSITY, AND MUSICAL EMOTION PERCEPTION

YURI BROZE, BRANDON T. PAUL, ERIN T. ALLEN &
KATHLEEN M. GUARNA
Ohio State University

THREE EXPERIMENTAL STUDIES SUGGEST THAT music with more musical voices (higher *voice multiplicity*) tends to be perceived more positively. In the first experiment, participants heard brief extracts from polyphonic keyboard works representing conditions of one, two, three, or four concurrent musical voices. Two basic emotions (happiness and sadness) and two social emotions (pride and loneliness) were rated on a continuous scale. Listeners rated excerpts with higher voice multiplicity as sounding more happy, less sad, less lonely, and more proud. Results from a second experiment indicate that this effect might extend to positive and negative emotions more generally. In a third experiment, participants were asked to count (denumerate) the number of musical voices in the same stimuli. Denumeration responses corresponded closely with ratings for both positive and negative emotions, suggesting that a single musical feature or percept might play a role in both. Possible roles for both symbolic and psychoacoustic musical features are discussed.

Received: January 8, 2013, accepted March 8, 2014.

Key words: musical emotion, musical voices, multiplicity, numerosity, denumerability, loneliness, Well-Tempered Clavier

IT HAS LONG BEEN SPECULATED THAT MUSIC is expressive of emotion in ways similar to the speaking voice (von Helmholtz, 1863; Kivy, 1980; Rousseau, 1782; Spencer, 1857). More recent empirical evidence has indicated that emotion cues in music and speech overlap considerably (Ilie & Thompson, 2006; Juslin & Laukka, 2003; Thompson, Marin, & Stewart, 2012), and many musical features are reliably associated with specific emotions (see Gabrielsson & Juslin, 2003; Gabrielsson & Lindström, 2010). Nonetheless, explanations by simple analogy or homology to speech cannot easily account for all musical situations. A great deal of music does not come from a single voice, and instead reflects the coordinated effort of many musicians performing simultaneously (see, e.g., Blacking, 1973; Merriam, 1964; Nettl, 2005). Indeed, unaccompanied melodies represent only a fraction of the music typically encountered, while most music heard today features multiple simultaneous musical parts, lines, or accompaniments. If musical emotions are perceived in part because musical and linguistic emotion cues coincide, then how do listeners perceive emotion in music that resembles more than one voice?

Relatively few psychological studies of musical texture and emotion appear to have been conducted, and those that have been published have produced conflicting results. Kastner and Crowder (1990) asked children aged three to twelve to respond to both monophonic and block-chord accompanied melodies by pointing to drawings of faces depicting positive and negative emotions. Overall, unaccompanied stimuli appeared to be perceived more positively than accompanied stimuli.[1] A similar correspondence was identified by Webster and Weir (2005) using four-part harmonizations. In their study, undergraduate participants consistently rated nonharmonized melodies as happier than harmonized melodies for both major and minor stimuli using a continuous scale.

Nonetheless, other research has produced precisely the opposite response pattern. In an attempt to replicate Kastner and Crowder (1990), Gregory, Worral, and Sarge (1996) asked both children and adults to identify accompanied and unaccompanied melodies as sounding either happy or sad. Unaccompanied melodies in fact sounded *less happy* than accompanied melodies, contradicting the earlier result. Similar results were obtained by McCulloch (1999), whose research also suggested that accompanied melodies are perceived more positively than unaccompanied melodies.

The lack of consensus regarding the emotional effects of textural density potentially reflects the different

---

[1] In fact, the authors analyzed their data using a 'hit or miss' score depending on whether the emotion response was congruent with major-positive/minor-negative associations. That unaccompanied melodies were perceived more positively can be seen by reflecting the minor-mode results in their Figure 3 about the *x*-axis (Kastner and Crowder, 1990, p. 199).

Music Perception, VOLUME 32, ISSUE 2, PP. 143–159, ISSN 0730-7829, ELECTRONIC ISSN 1533-8312. DOI: 10.1525/MP.2014.32.2.143

versions of dense textures used in these studies. Kastner and Crowder (1990) employed block chords played on the piano, and Webster and Weir (2005) appear to have used basic chorale-style harmonization. McCulloch (1999) used accompaniments taken from published songbooks, and Gregory et al. (1996) appears to have done the same. One could imagine several musical features which might have differed between these harmonized or accompanied dense textures. The present study will diverge from these previous approaches by using contrapuntal polyphonic stimuli of four different levels of textural density. Hopefully, our results will provide a helpful third perspective.

Approaches to Musical Texture

When describing different textures, musicians often refer to relationships between musical parts, or voices. In this context, a 'musical voice' does not necessarily need to be sung, but instead can refer generically to any horizontal musical line. As a result, the relationship between the number of musical voices and the number of performers is not always one-to-one: many complex arrangements are possible (Cambouropoulos, 2008). A *monophonic* texture could arise both from a single person chanting or from several instrumentalists performing in unison. Similarly, *polyphonic* music with multiple musical voices could be performed by a single musician using a polyphonic instrument.[2] While polyphony in the broadest sense can refer to any music with more than one simultaneous note, it also can be used in a more specific sense to designate contrapuntal textures with a high degree of independent linear motion (such as occurs in canons, fugues, and Renaissance motets).

In addition to monophony and polyphony, several other types of texture have been identified by music theorists. *Homophonic* music employs a texture based on chords: many different notes appear at once, but their rhythms are aligned, as in traditional chorale settings (Clendinning & Marvin, 2005; Laitz, 2012). Some have identified 'heterophony' as a distinct texture in which a single melody is simultaneously presented in multiple variations (Cooper, 1981). Berry (1976) adopts an open-ended interpretation of musical texture, accommodating many different intervallic, rhythmic, and melodic components. Other theorists, such as Smith Brindle (1966) and Rowell (1983) have focused on the global sound quality of the texture. Dimensional approaches have also appeared: Huron (1989a) describes monophony, polyphony, homophony, and heterophony as four regions in a two-dimensional texture space, representing different degrees of semblant motion and onset synchrony.

The present study investigates the perception of musical emotion in terms of musical *voice multiplicity*: the number of musical parts or voices simultaneously present in a texture. Because our primary goal was to measure a purely musical effect (and not one which would actually reflect the number of musicians performing), we used stimuli performed by a single person on a polyphonic instrument. A straightforward definition of musical voice was adopted: the number of notated musical lines specified by the composer of contrapuntal works.

Because the voice multiplicity of a piece of music might implicitly or explicitly designate certain social settings, we conjectured that listeners would be more likely to perceive musical emotions corresponding to the musical texture's social connotation. We were inspired by recent exploratory research on musical affect: Albrecht's (2012) complete surface-level affective analysis of the slow movement from Beethoven's Sonata No. 8, Op. 13 ("Pathétique"). In Albrecht's 'progressive exposure' paradigm, over two hundred listeners were asked to judge the emotional content of five-second excerpts representing the entire duration of the movement. One of the rated emotions, loneliness, appeared to show sensitivity to the number of musical voices present (see Figure 1).

Specifically, we hypothesized that thin textures should sound more lonely than thick textures. Note that this prediction runs somewhat contrary to Webster and Weir (2005) and Kastner and Crowder (1990), who reported that monophony tends to be associated with increased positive affect over harmonized (multivoice) textures.

In our first experiment, participants rated brief musical excerpts for four emotions, including perceived loneliness. We predicted that music with fewer musical voices would tend to sound lonelier than music with many musical voices. A second experiment tested whether voice multiplicity effects generalize to emotional valence and sociality. Finally, a third experiment allowed subjects to generate their own emotion labels and directly measured their ability to denumerate the voices in the specific stimuli used. To anticipate our results, listeners appear to associate increasing polyphonic voice multiplicity with increasing emotional positivity. Moreover, ratings for perceived emotion will bear strong resemblance to the perceived number of musical voices present, suggesting that perception of

---

[2] A musical voice might even refer to imagined melodies that are not fully articulated, as occurs in pseudopolyphony (Davis, 2006).

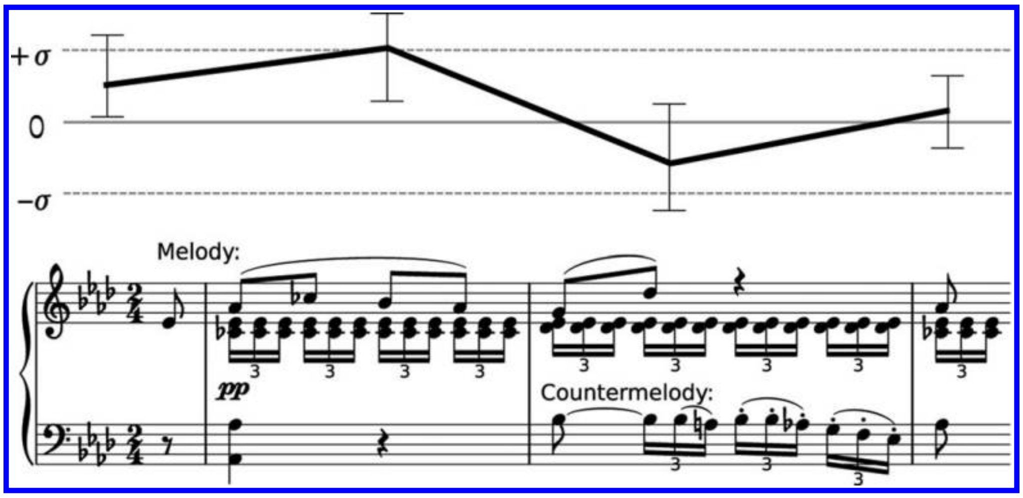

FIGURE 1. Loneliness ratings for the minor-mode episode's first two measures are given as z-scores normalized per person. Whiskers indicate the 75th and 25th percentiles of all rating z-scores. Loneliness ratings begin rather high when the melody is initially heard, but drop sharply after an additional voice enters the texture. The present investigation. Figure adapted from Albrecht (2012, Appendix A).

both emotion and voice numerosity might rely on the same underlying cues.

Experiment 1: Perception of Musical Loneliness

Our conjecture that textures with fewer voices should sound more lonely implies that listeners form some perceptual or cognitive representation of the number of voices present. We shall refer to this proposed subjective representation as the perceived *voice numerosity* of a piece of music.[3] The distinction between voice multiplicity and voice numerosity reflects the division between external acoustic sound sources and internal auditory percepts, similar to that between amplitude and loudness, or frequency and pitch. Note that music's voice numerosity need not match its nominal voice multiplicity: perceptual errors could be made. In this case, perceived musical loneliness would be expected to reflect subjective voice numerosity, including any perceptual errors that might have arisen.

As a subjective psychological construct, voice numerosity cannot be directly measured. However, an indirect measure is possible: performance on a voice *denumeration* task, in which a listener reports the number of voices present in a polyphonic texture. Voice denumeration appears to be subject to certain limitations. While listeners are typically accurate in identifying one-voice and two-voice textures, they increasingly underreport the voices present in denser textures (Huron, 1989b; Parncutt, 1993). In fact, Huron's listeners reported four-voice textures as having only three voices more often than they answered correctly (see Table 2). It would therefore appear that as polyphonic textures grow denser, voice multiplicity levels become less perceptually distinct.[4] Presuming these voice denumeration results reflect perceptual voice numerosity, we additionally predicted that perceived loneliness should change more between one- and two-voice textures than between three- and four-voice textures.

STIMULI

Fugal compositions were chosen because they feature polyphonic textures with clearly defined voice multiplicities. Fugues typically begin with a single monophonic melody: the fugue's *subject*. Additional voices then accumulate, with each new entry repeating the subject melody. By excerpting from fugue expositions, sets of stimuli matched for melodic content, tempo, and other composition-specific features can easily be generated. Recordings of J. S. Bach's *Well-Tempered Clavier* by a professional harpsichordist playing a 1624 *clavecin* were used (Bach, 1722; Verlet, 1993). Because the harpsichord's plucking mechanism results in relatively uniform

---

[3] This use of 'numerosity' roughly aligns with its use in the study of numerical cognition (Cheatham & White, 1954; Dehaene, 1997; ten Hoopen & Vos, 1979; White, 1963).

[4] As Huron (1989b) points out, this is reminiscent of the '*un, deux, trois, beaucoup*' pattern of responses identified by Descoeudres (1921) in her studies of numerical cognition in children. However, because listeners often misidentified four voices as only three, perhaps 'one, two, many' would better describe polyphonic voice numerosity.

intensity, articulation, and timbre across its range, performers have limited ability to highlight certain musical voices over others.

Because loneliness is a negatively valenced emotion, we expected it would be most perceived in minor-mode pieces of music. Hence, all minor-mode fugues of at least four voices were identified. For each, the beginning and ending of each statement of the fugal subject was marked, and 5 s excerpts from the center of each multiplicity region were generated, including 200 ms logarithmic fade-ins and fade-outs. The G minor fugue was excluded because its exposition does not include stepwise accumulation of voices, and fugues without 5 contiguous seconds for each multiplicity were also excluded. This left seven minor-mode fugues with four isotextural multiplicity conditions each. Two major fugues from the same recording were additionally sampled in order to increase the modal diversity of the stimuli. Ratings for major-mode stimuli were not analyzed in the first experiment.

PROCEDURE

Twenty-eight Ohio State University School of Music undergraduates enrolled in sophomore-level aural skills classes participated for course credit (17 female, 11 male, aged 19–21). Participants were asked to provide emotion ratings for musical excerpts played through free-field speakers at a comfortable level. All trials took place in an Industrial Acoustics Corporation sound-attenuation chamber.

In addition to loneliness, three other emotions were rated to disguise the target emotion, as well as to provide data for exploratory analysis. Subjects rated the prototypical negative and positive emotions of 'sadness' and 'happiness'. Additionally, 'pride' was included as a socially-relevant positive emotion. Participants were instructed to rate the emotions they perceived in the music (not necessarily the emotions they experienced themselves).

Each computer-administered trial presented the prompt: "How [emotion] does this music sound?" Responses were collected using a continuous slider interface labeled only on the left and right as "not at all [emotion]" (scored as 1.0) and "very [emotion]" (scored as 7.0). After each rating, the slider was reset to the center position (4.0). After five practice trials, participants rated all 144 stimulus-emotion combinations in random order. A second block of 72 randomly-selected stimulus-emotion pairs immediately followed, used to gauge the reliability of emotion ratings. Post-experimental interviews were conducted to check for demand characteristics and explore strategies used.
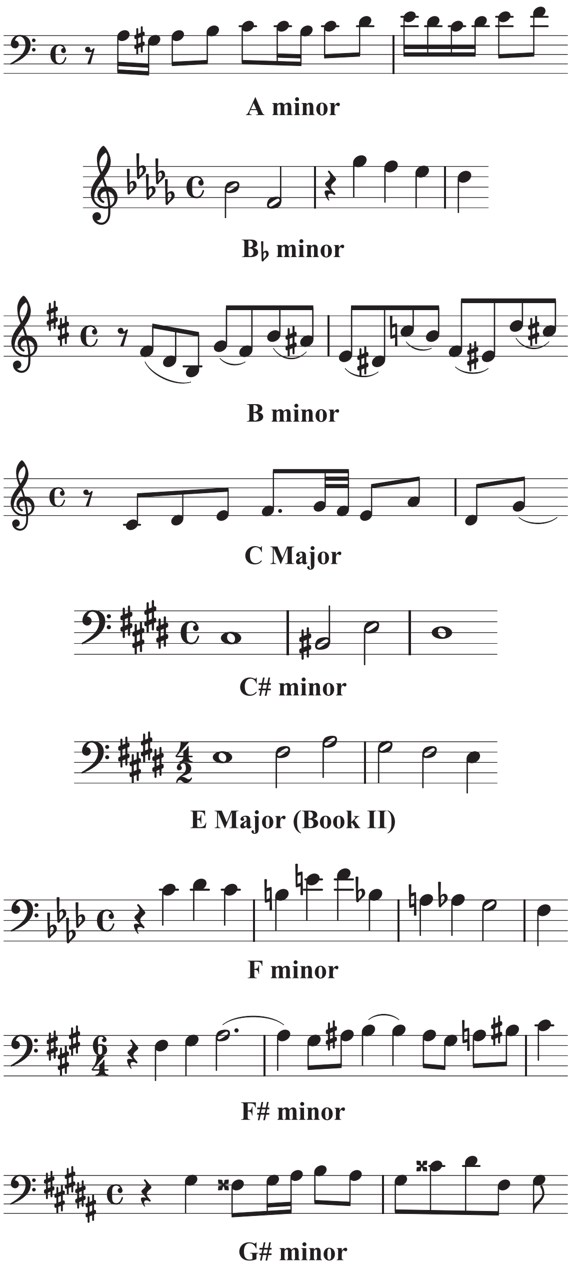
FIGURE 2. From each of nine fugues, four 5 s stimuli were generated; one for each voice multiplicity condition. Fugues are from J.S. Bach's *Well-Tempered Clavier* Book I, except for the E major fugue which was taken from Book II.

RESULTS

Across all repeated subject-stimulus ratings for loneliness, the test-retest correlation was $r_s = .65$, and the average difference in rating was unbiased, with a 95% confidence interval of (−0.15, 0.14). Duplicate measurements were therefore averaged prior to further analysis in order to increase analytic precision.

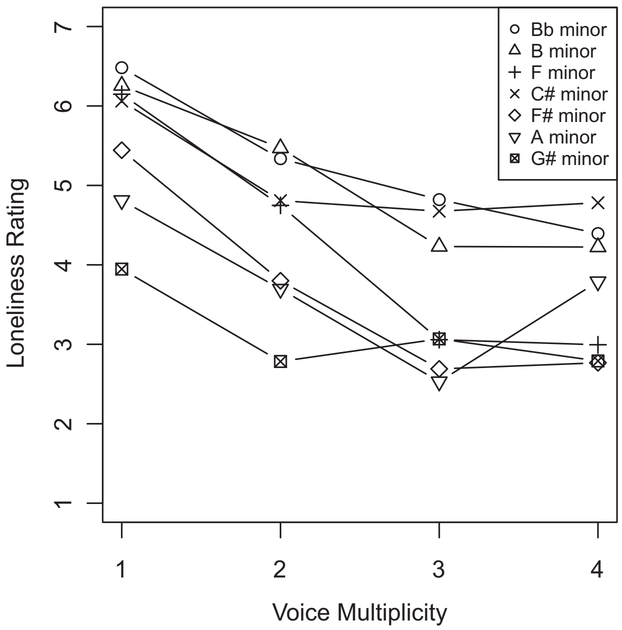

FIGURE 3. Median loneliness ratings for stimuli from each fugue used in experiment one are plotted. Stimuli were 5 s excerpts of fugal expositions from Bach's *Well-Tempered Clavier*. Keys are given to identify the fugue, and do not indicate a generalization of effects to key. Perceived loneliness drops as voice multiplicity increases, with listeners rating one-voice stimuli as 1.89 points lonelier than four-voice stimuli on average, $t(27) = 8.71$, $p < .001$. Moreover, this effect leveled off as voice multiplicity increased: loneliness ratings dropped more between one- and two-voice conditions than between three- and four-voice conditions, $t(27) = 5.21$, $p < .001$.

Figure 3 shows median loneliness ratings for each minor fugue in the study, as affected by the excerpt's voice multiplicity. In general, loneliness ratings decrease as voice multiplicity rises, regardless of fugue. On average, listeners rated a given fugue's one-voice condition as 1.89 points lonelier on a 7-point scale than its four-voice presentation, $t(27) = 8.71$, $p < .001$, consistent with the first hypothesis. A second hypothesis predicted that loneliness ratings would decline the least between three- and four-voice conditions, because these multiplicities are phenomenological similar. On average, listener ratings dropped 0.99 points more between one- and two-voice conditions than between three- and four-voice conditions, $t(27) = 5.21$, $p < .001$. This result is consistent with the idea that loneliness ratings and voice denumeration are both mediated by perceptual voice numerosity.

Other emotions. The effect of voice multiplicity on perceived happiness, sadness, and pride were also explored (Figure 4). The positively valenced emotions happiness and pride both exhibited increased ratings for higher voice multiplicity, while ratings for the negatively valenced emotions of sadness and loneliness showed the opposite effect. It would appear that the positive/negative valence of an emotion dictated the direction of the relationship.

Additionally, the effect of voice multiplicity on ratings for the social emotions (loneliness and pride) appears stronger than for the nonsocial emotions (sadness and

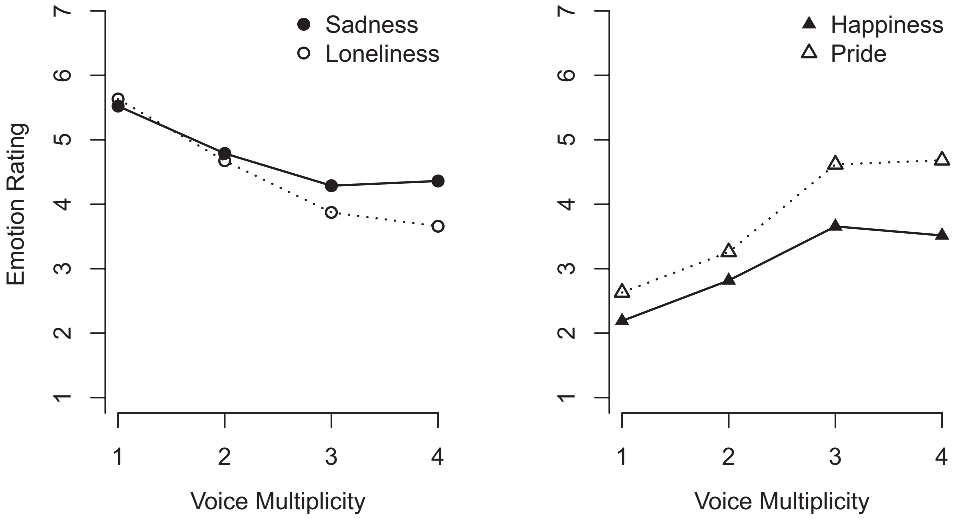

FIGURE 4. Median emotion ratings for negatively and positively valenced emotion pairs from Experiment 1. Filled points (●, ▲) represent the nonsocial emotions sadness and happiness, while open points (○, △) represent the social emotions loneliness and pride.

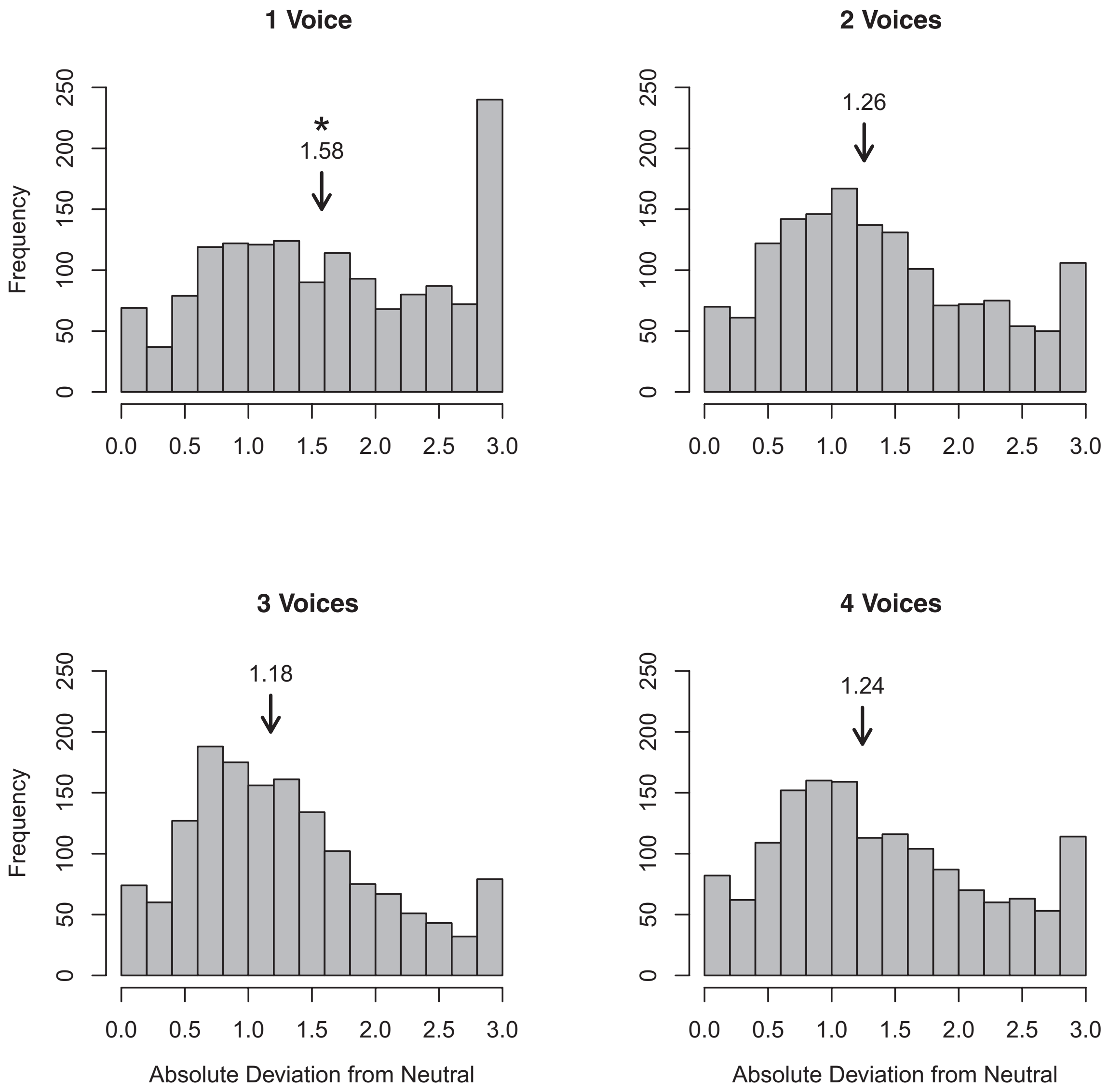

FIGURE 5. Histograms summarizing a post hoc analysis of emotion rating strength in Experiment 1. The absolute deviation of each slider-based emotion rating from the center position for different voice multiplicities was calculated for all four emotions. Medians are indicated with arrows (↓). Pairwise Mood median tests indicated that the median emotion rating for the 1-voice stimuli was different from all three others, $p < .0001$, indicated by a star (*).

happiness). A post hoc test revealed that emotion rating differences between one- and four-voice textures was 0.65 points greater for loneliness than for sadness, $t(27) = -3.09$, $p = .003$. The same result obtained for positively valenced emotions: pride ratings changed 0.46 points more than happiness ratings, $t(27) = 2.55$, $p = .009$.[5]

Strong ratings of musical emotion. The preceding analysis treats ratings as distances from the scale's 'zero' point ('not at all [emotion]'). However, one could also consider ratings as directed deviations from the slider's center point. Viewed this way, emotion ratings appear more *neutral* when more voices are present, and stronger when fewer voices are present (see Figure 4).

To explore this idea, the strength of each emotion rating was calculated by measuring its absolute deviation from the scale's midpoint (4.0). Histograms of absolute rating strength for each voice multiplicity are given in Figure 5. Results of six post hoc Mood median tests suggest that the monophonic textures' median rating strength was higher than each of the other three

---

[5] Note, however, that the second experiment failed to replicate this finding.

($p < .0001$ for every comparison), while other comparisons were statistically nonsignificant. In other words, emotion ratings in experiment one appear to have been particularly strong for monophonic stimuli, whereas rating strength for multiplicities of two, three, and four were statistically indistinguishable.

DISCUSSION

Our results are consistent with the idea that voice multiplicity is negatively associated with perceived loneliness. Post-experimental interviews corroborated this finding: many participants reported that textural density informed some emotion judgments. This effect appears to correspond with perceptual limitations on voice denumerability: voice multiplicity effects arose predominantly between one-, two-, and three-voice conditions. This raises the possibility that perceptual voice numerosity might mediate musical emotion perception. Additionally, post hoc analysis of rating strength indicated that monophonic textures result in qualitatively different responses from other multiplicitly levels.

Second, the results suggest that voice multiplicity effects might generalize to many emotions based on their valence (positive or negative), their sociality (social or nonsocial), or both. However, in the present experiment each participant rated all four emotions, with several expressing awareness of the 2 × 2 design in post-experimental interviews. To address this possible demand characteristic, our second experiment used a between-subjects design to test whether social emotion ratings are more strongly affected by voice multiplicity than nonsocial emotion ratings.

Experiment 2: Generalization to Emotion Valence and Sociality

Although the above results suggest that voice multiplicity might influence perceived emotion ratings in interaction with emotional valence and sociality, the experimental design introduced a possible demand characteristic. This second experiment provides an explicit test for these effects. A between-groups design was used with regard to valence, and within each valence group, three social and three nonsocial emotions were chosen.[6] In all, three predictions were tested: First, positive emotion ratings should increase, and negative emotion ratings should decrease as voice multiplicity rises. Second, social emotions should exhibit larger-magnitude voice multiplicity effects than nonsocial emotions. Third, we hoped to reproduce the result that monophonic stimuli evoke stronger responses than any other multiplicity level.

Categorizing emotions into 'social' and 'nonsocial' types draws on experimental evidence from developmental studies (Kagan, 1981; Stipek, Gralinski, & Kopp, 1990), and more recently, functional neuroimaging (e.g., Burnett, Bird, Moll, Frith, & Blakemore, 2009). Experiencing certain emotions—such as pride, shame, guilt, embarrassment, and envy—depends on the capacity to model the mental states of oneself and others. These emotions have been variously referred to as social emotions (Minzenberg, Poole, & Vinogradov, 2006; Teroni & Deonna, 2008), self-conscious emotions (Lewis, Sullivan, Stanger, & Weiss, 1989; Tracy, Robins, & Tangney, 2007), or moral emotions (Eisenberg, 2000; Tangney, Stuewig, & Mashek, 2007). Social emotions often overlap with so-called 'secondary emotions,' distinguishable from the 'primary' or 'basic emotions' which appear early in life (Lewis & Michalson, 1983), and have cross-culturally recognizable facial expressions (Ekman, 1973). Both loneliness (Cacioppo & Patrick, 2008; Weiss, 1975) and pride (Tracy & Robins, 2007; Williams & DeSteno, 2008) have been described as social emotions.

Although we interpreted the preceding results as suggesting an emotion sociality effect, emotions like loneliness and pride could also be distinguished from sadness and happiness by other factors, such as categorical prototypicality (Lako, 1987; Rosch, 1973, 1975), or familiarity (Bornstein, 1989; Zajonc, 1968). Experiment 2 uses both as covariates.

STIMULI AND PROCEDURE

In order to choose a well-stratified collection of social and nonsocial emotions of both positive and negative valence, we made use of the analysis of Shaver, Schwartz, Kirson, and O'Connor (1987), who clustered 135 emotions into hierarchical categories using results of a sorting task. Six positive and six negative emotions (split by sociality) were chosen such that no two emotions fell in the same subcategory (see Table 1). Prototypicality covariates were taken from Shaver et al.'s participant's subjective ratings. To estimate each emotion word's familiarity, its frequency ranking in the Corpus of Contemporary American English was found (Davies, 2008—). These two covariates would be used for post hoc exploratory analysis.

Sophomore-level music students aged 19–21 participated for Aural Skills course credit. Participants were alternately assigned to the positive-emotion ($N = 18$) or the negative-emotion group ($N = 19$) in the order in

---

[6] Given the number of emotions rated, it was expected that the distinction between social and nonsocial emotion types would not be immediately evident to subjects.

TABLE 1 *Emotions Used in Experiment Two.*

Valence Group	Sociality	Emotion	Prototypicality	Frequency
Positive	Social	Pride	3.14	13805
		Love	3.94	153466
		Compassion	3.62	5461
	Nonsocial	Happiness	3.77	8177
		Contentment	2.92	796
		Excitement	3.51	9231
Negative	Social	Shame	3.44	9319
		Loneliness	3.41	3121
		Envy	3.58	3355
	Nonsocial	Sadness	3.68	4734
		Fear	3.83	49410
		Disgust	3.42	2879

Note. Emotions rated in the second experiment on emotion valence and sociality. Prototypicality scores are average ratings on a 4-point scale provided by 112 university students (Shaver et al., 1987). Word frequency reflects the number of occurrences in a corpus of 450 million words used in contemporary American English (Davies, 2008–).

which they arrived. Using identical slider interfaces as in the first study, the positive-emotion and negative-emotion groups rated identical stimuli for the emotions given in Table 1, in randomized order. Because the major-mode and minor-mode excerpts did not appear to elicit different response patterns, results for all 36 stimuli were analyzed together (7 minor fugues and 2 major fugues with four voice multiplicity conditions).

RESULTS AND DISCUSSION

Results for the positive and negative emotion groups are compared in Figure 6. It is clear that positive and negative emotion ratings responded in opposite directions to increasing polyphonic voice multiplicity. The mean difference in emotion rating moving from monophonic to four-voice music was −0.83 for the negative emotion group, $t(18) = 6.10$, $p < .001$, and +1.44, and for the positive emotion group, $t(17) = 12.2$, $p < .001$. In fact, all twelve emotions responded in the expected direction, with positive emotion ratings increasing with voice multiplicity and negative emotion ratings decreasing. In short, it seems that thinner polyphonic textures are associated with negative emotional valence and thicker textures with positive emotional valence. Once again, the effect appeared to taper off as multiplicity rose.

To test for an effect of emotion sociality, two-way mixed models were fit to the responses of each the positive and negative emotion groups, using multiplicity and sociality to predict emotion ratings, blocking by both subject and fugue.[7] The hypothesis that social emotions would respond more strongly than nonsocial emotions would predict that an interaction between voice multiplicity and sociality should be present. To control for multiple tests, p values below .01 were treated as statistically significant, for a familywise error rate under .06.

Main effects were identified for voice multiplicity (considered linearly) in both the negative and positive groups, with higher multiplicity associated with lower negative emotion ratings, $\chi^2(1) = 183$, $p < .001$, and higher positive emotion ratings, $\chi^2(1) = 556$, $p < .001$. Additionally, a main effect was found for emotion sociality in the positive emotion group, $\chi^2(1) = 129$, $p < .001$; social positive emotions got higher ratings than nonsocial positive emotions overall. However, no significant interaction effect between voice multiplicity and emotion sociality was found for either emotion valence group, failing to support the hypothesis that social emotion perception ought to depend on multiplicity more strongly than nonsocial emotion perception; $\chi^2(1) = 0.18$ for negative emotions and $\chi^2(1) = 0.91$ for positive emotions.

A post hoc reanalysis using the assembled emotion prototypicality and frequency covariates instead of sociality indicated that emotion ratings did not depend on prototypicality, but might have depended on an emotion term's frequency of occurrence. Specifically, positive emotions were rated higher when the word was more frequent, $\chi^2(1) = 10.1$, $p = .002$, and negative emotions were rated lower, $\chi^2(1) = 21.4$, $p <. 001$. This potentially reflects a version of the 'mere exposure' effect, wherein more frequently-encountered stimuli are perceived more positively (Zajonc, 1968). This would indicate that emotion ratings at least partially reflect listener preferences and attitudes toward the music or

---

[7] Although the slider interface introduced ceiling and floor effects, these would be expected to reduce statistical power, increasing false negatives, not false positives. Model diagnostics indicated that residuals were well-behaved.

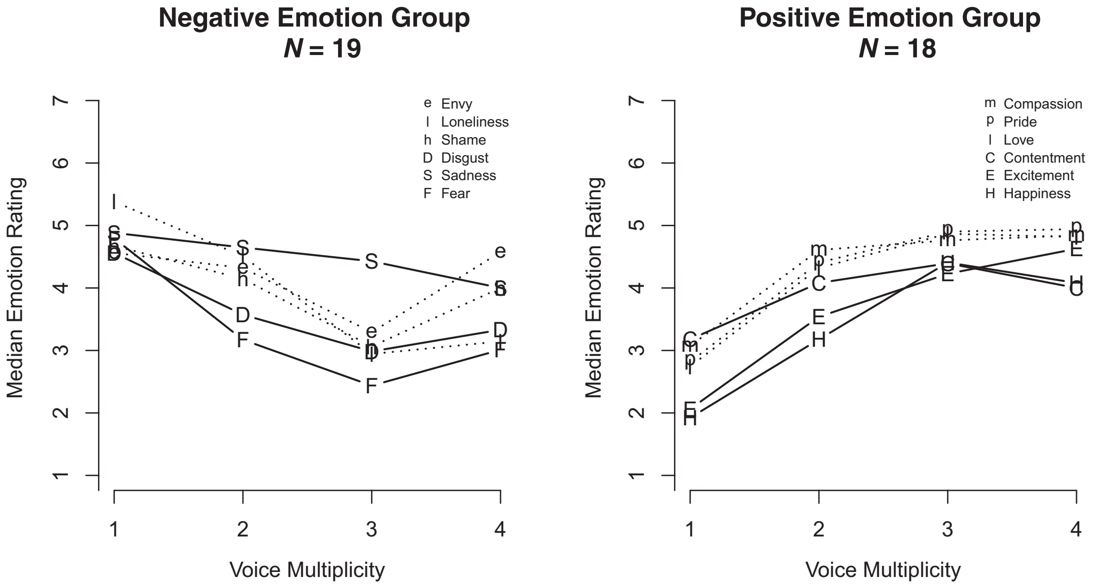

FIGURE 6. Median emotion ratings per emotion are given for the positive ($N = 19$) and negative ($N = 18$) emotion groups in the second experiment. As predicted, positive emotion ratings increased with increasing voice multiplicity while negative emotion ratings decreased. Using ANOVA analyses, main effects were identified for voice multiplicity in both groups. However, emotion sociality did not appear to change the magnitude of this effect, contrary to the experimental hypothesis.

the task demands, not only the music itself. Because of the high familiarity of the positive social emotions (particularly 'love'), the observed main effect for sociality might be spurious. Although the absolute ratings appeared to be affected by emotion word frequency, neither covariate seems to have impacted the absolute strength of emotion ratings.

Finally, monophonic stimuli again resulted in extreme emotion ratings more frequently than any other multiplicity level, repeating the finding from the first experiment, logistic $z = 7.76$, $p < .001$.

Experiment 3: Emotion Rating Strength and Voice Denumerability

Results of the first two experiments indicate that as polyphonic textures thicken, negative emotion perception decreases while positive emotion perception increases. Additionally, the effect appears to taper off in a manner similar to limitations in voice denumerability, with 3 voice and 4 voice textures exhibiting smaller differences in emotion perception than 1 and 2 voice textures. This is consistent with the idea that perceptual voice numerosity mediates both responses.

However, this interpretation relies on denumeration data from separate studies using different stimuli and procedures: Huron (1989b) and Parncutt (1993). To provide internal validity, experiment three endeavors to replicate Huron's polyphonic voice denumeration results using the current protocol. Additionally, we will attempt to bolster the external validity of emotion perception findings by using only the most expressive subset of stimuli and by allowing listeners to freely specify which emotions they perceive.

In order to make more refined predictions regarding voice denumerability, it is worthwhile to imagine perceptual or cognitive processes that could produce a subjective sense of voice numerosity or otherwise determine denumeration responses. First, listeners could estimate the number of voices, making approximate, nonsymbolic judgments of magnitude based on some feature of the sound, such as event density. Second, listeners might be able to quickly identify certain voice multiplicity levels without counting, in a process that has been called 'subitizing' in the context of vision and numeric cognition research (Jevons, 1871; Kaufman, Lord, Reese, & Volkmann, 1949).[8] Subitizing would seem most likely in a musical context for small voice

---

[8] There is some debate as to whether subitizing in fact reflects a faculty distinct from counting. For present purposes, we treat it separately based on the rapidity with which it appears to operate.

multiplicities such as one or two. Third, listeners could deliberately count the number of voices by shifting their attention between perceived auditory streams (Bregman, 1990; Bregman & Campbell, 1971). Counting streams could be facilitated by recognizing voice entries and exits, but need not rely on them exclusively.

Note that these processes differ as to the response profile they would be expected to produce with regard to central tendency (accuracy) and statistical dispersion (precision). For example, subitizing ought to have both high accuracy and high precision, as it represents a direct apprehension of the voice multiplicity. By contrast, estimation would be expected to be somewhat imprecise, and potentially systematically biased as well (e.g., Tversky & Kahneman, 1974). Counting processes could result in several response patterns, possibly depending on a listener's streaming capacity or their working memory. In general, one might expect counting responses to be reasonably precise, but potentially biased in one direction or another, depending on whether voices are actively entering or exiting the music.

Because they listened to music as it unfolded, listeners in Huron's denumeration study would presumably have had all three processes available; however, the current stimuli might not offer such flexibility. Huron's listeners would have been able to identify entries and exits, but the stimuli used here are isotextural, containing no voice entries or exits by design. The excerpts' short length could create further problems, as stream segregation appears to be cumulative (see Bregman, 1978): listeners tend to hear one stream at first before subsequently resolving more. In all, because our stimuli are both short and isotextural, counting would not be expected to be a successful strategy in the present denumeration task.

In addition to replicating Huron's results, this third experiment tests the hypothesis that voice denumeration and emotion ratings will follow similar response patterns with regard to voice multiplicity. Additionally, we again tested whether monophonic stimuli would be more likely to evoke extreme emotion ratings than multi-voice stimuli.

PROCEDURE AND STIMULI

Procedures were similar to those of the first two experiments, but attempted to minimize a possible design artifact due to nonexpressive musical excerpts: If listeners do not hear the music as emotionally expressive of the presented emotion (or any emotion for that matter), then their emotion ratings might have limited meaning. In order to ensure the fugues employed were sufficiently expressive, all $36 \times 12 = 432$ stimulus-emotion combinations from Experiment 2 were ranked in terms of their median rating's absolute deviation from the middle point of the scale. The rank sum of these deviations across each fugue's four multiplicity levels for all emotions was used as a measure of the fugue's expressivity. Of the nine fugues, the B minor, E major, C major, and B♭ minor fugues were selected as the most expressive.

Participants were undergraduates enrolled in sophomore-level aural skills classes and participated for course credit ($N = 24$, 10 females, 14 males, aged 19–21). The study was divided into three parts: emotion generation, emotion rating, and denumeration. In the first part, subjects listened to four excerpts as many times as they wished. For each, subjects generated three emotion labels completing the sentence, "I hear _____ in the music." They subsequently chose the one emotion label that fit the excerpt the best. The four stimuli used to generate emotions represented each of the four fugues and each of the four voice multiplicity levels; each of the 24 subjects listened to one of the possible subsets of stimuli satisfying this constraint in random order.

The second part of the experiment again employed the 1.0-to-7.0 continuous slider paradigm to rate emotion perception, using the four listener-selected emotion labels in combination with the 16 different musical excerpts for 64 total randomized emotion-excerpt trials. In the third part, subjects denumerated the voices in excerpts by typing a single digit into an empty text field. All sixteen excerpts were presented twice, in two randomized blocks. Participants then completed the Ollen Musical Sophistication Index battery (Ollen, 2006).

Volunteered emotions were classified as positive, negative, or neutral/don't know by an independent group of six raters. Emotions for which five of six raters agreed were grouped into valence categories for analysis.

RESULTS AND DISCUSSION

Denumeration. Voice denumeration responses (Table 2) ranged from one to six voices, and the test-retest correlation between the two denumeration blocks was. For each multiplicity level tested, listeners' most common voice denumerations matched those of Huron (1989b), with the mode response in both studies being 1-2-3-3 for the first four multiplicities. Average denumeration responses are quite concordant as well, with underreporting being much more common than overreporting (see Figure 7, top panel). Listeners with higher OMSI scores made fewer errors overall (logistic $z = 2.10$, $p = .036$).

We additionally tested whether the form of denumeration error types was concordant with Huron's

TABLE 2. *Voice Denumeration Confusion Matrices.*

		Denumeration Response					
	Actual Multiplicity	1	2	3	4	5	6
(a) Huron (1989b)							
	One	**309**	6	15	0	3	0
	Two	19	**1368**	63	1	0	0
	Three	0	127	**1121**	12	0	0
	Four	0	77	**961**	608	14	0
	Five	0	22	165	**466**	399	4
(b) Experiment 3							
	One	**185**	6	0	0	0	0
	Two	2	**151**	37	2	0	0
	Three	0	87	**95**	9	0	0
	Four	0	43	**82**	64	1	1

Note. Results from Huron1989 are continuous denumeration responses by five musicians listening to J. S. Bach's "Saint Anne" organ fugue. Results from Experiment 3 are from 24 undergraduates enrolled in a music course, and correspond to isolated 5 s excerpts of harpsichord music. Mode responses per multiplicity level are in bold text. In both cases, underreporting predominated, with the most common responses following a 1-2-3-3 pattern for the first four multiplicities.

results. The confusion matrices from each study were normalized by number of trials per multiplicity level, and denumerations greater than five were truncated. The probabilities of each incorrect response type in the confusion matrices (i.e., non-diagonal entries) were tested for statistical dependence using Kendall's tau. The patterns of erroneous responses were highly correlated, $\tau(16) = .71$, $p < .001$, reflecting rank-order correspondence of both bias and dispersion effects between the studies.

Despite this rank-order correlation, error rates in the present study were substantially higher for 2-voice and 3-voice stimuli than in Huron's previous results (Figure 7, bottom panel). This could reasonably be attributed to the aforementioned differences between the two tasks: while Huron's listeners would have been able to count voice entry and exits over a more typical listening experience, our subjects had only five seconds of music with no entries or exits.[9] The uniformly high precision and accuracy for monophonic stimuli suggest that instant recognition (subitizing) might have occurred in both studies for one-voice textures. While Huron's listeners would have potentially been able to apply counting techniques to detect the entry of a second or third voice, our listeners were presumably much more likely to use estimation because the stimuli were isotextural. Finally, in both case listeners appeared to resort to estimation processes when denumerating voices in four-part textures, resulting in very high error rates.

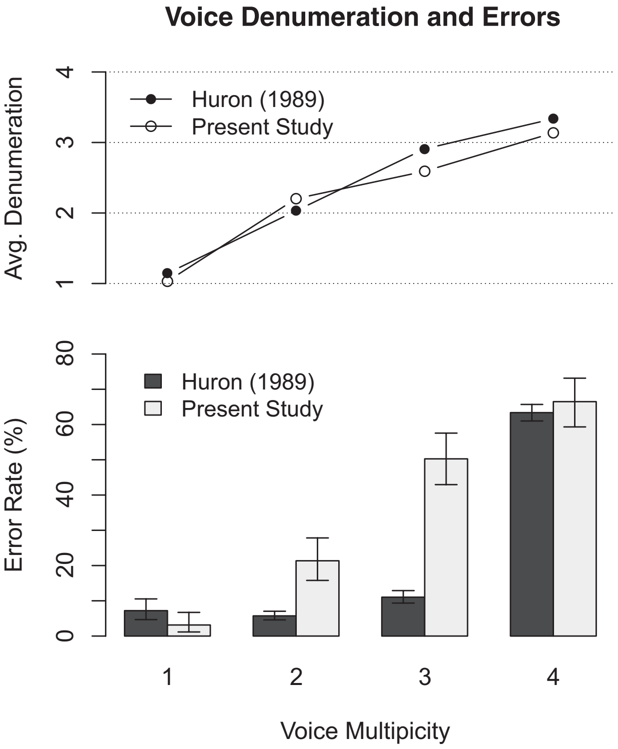

FIGURE 7. Comparison of average voice denumeration response and error rates obtained in Huron (1989) and Experiment 3. Huron's results are from five expert musicians who listened to J. S. Bach's "Saint Anne" organ fugue and continuously reported the voices present. In the present study, 24 undergraduate music students responded to five-second harpsichord stimuli. In both cases, underreporting was more common than overreporting, and the central tendancy of responses was similar. Error rates were much higher in the present study for two-voice and three-voice textures. Error rate whiskers are raw 95% binomial confidence intervals.

---

[9] Additionally, Huron's participants were five expert musicians, while our participants were undergraduate music students. Given that higher OMSI scores correlate with denumeration performance, this disparity could also be attributed to differences in skill.

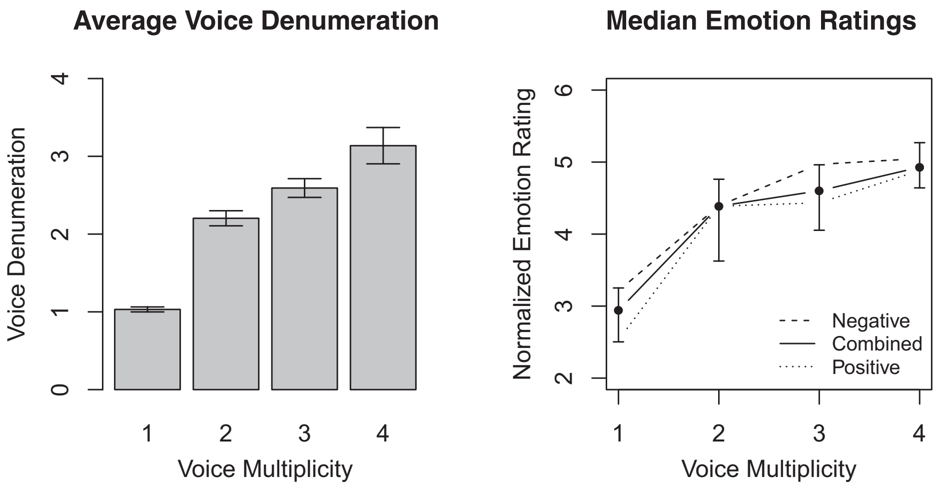

FIGURE 8. Voice denumeration compared with median ratings for negative and positive emotions in Experiment 3. Emotions were classified by an independent panel of judges; negative emotion ratings are plotted on an inverted scale to facilitate comparison. Both voice denumeration and emotion ratings tend toward similar response patterns, consistent with the idea that perceptual voice numerosity contributes to both. Subject-wise average voice denumerations and average emotion ratings per multiplicity level were strongly correlated, for both positive emotions, $r = .58$, and negative emotions, $r = -.56$. Denumeration whiskers are raw 95% $t(23)$ confidence intervals; emotion rating whiskers are 95% binomial confidence intervals around the median.

Emotion perception. Positive emotion ratings increased with voice multiplicity while negative emotion ratings decreased, replicating the emotion valence effect identified in Experiment 2. Median positive and negative emotion ratings are displayed in Figure 8, with the negative emotion rating scale inverted to facilitate direct comparison. Qualitatively, it would appear that median emotion ratings do indeed follow a similar pattern to average voice denumeration outcomes. Both positive emotion ratings ($r = .58$) and negative emotions ($r = -.56$) were strongly correlated with average voice denumerations. As an inferential test, a linear model was used to predict normalized emotion ratings using a subject's first set of denumeration responses, blocking by subject and fugue. A main effect for voice denumeration was identified, $\chi^2(5) = 120.1$, $p < .0001$, consistent with the idea that the two tasks depend on the same information.

One could describe the similarity of denumeration and emotion responses as appearing biphasic: both seem to respond approximately linearly for voice multiplicities of two and higher, but monophonic stimuli are qualitatively different. This could correspond to estimation processes and subitizing, respectively.

Emotion strength and monophonicity. Our final hypothesis predicted that the proportion of strong emotion ratings of 1.0 or 7.0 would be highest for monophonic stimuli. However, the probability that a listener would make a strong rating had no statistically significant relationship with voice multiplicity in experiment three, failing to support the hypothesis (logistic $z = -1.282$, $p = .2$). In short, monophonic stimuli were associated with stronger emotion ratings when emotions were experimenter-provided, but not when they were listener-generated.

One might suppose that when participants were provided with emotion words in the first two experiments, these terms were understood as *personal emotions* describing affective experience of an individual. If so, then perhaps monophonic stimuli were rated more strongly simply because they resemble a single person's voice. In this view, polyphonic textures might have resembled multiple agents each capable of a different emotional state, making it difficult to read any single emotion. An alternative account would be that monophonic stimuli simply have fewer relationships present, making emotion perception more straightforward. There might also be attentional influences leading to intervoice interactions: unattended musical voices appear to influence the perception of attended musical voices (Davison & Banks, 2003; Fujioka, Trainor, Ross, Kakigi, & Pantev, 2005). Regardless of the reason, it would seem that experimenter-provided emotion words might have led participants to seek voice-like *melodies* by which to judge emotional content.

By contrast, the third experiment requested that listeners provide their own emotion labels, prompted by musical stimuli. These emotions might have been understood as *musical emotions* with specific musical meanings and referents. Thus, participants in experiment three might have been making ratings based on an excerpt's musical similarity to the one that initially evoked the emotion instead of to a more typical understanding of individually experienced emotions. Because emotion terms were generated for all four multiplicities, it is understandable why monophonic stimuli would not always evoke the strongest responses.

General Discussion

Overall, our results are consistent with the idea that textural density exerts a general influence on the perception of musical emotion, with thicker polyphonic textures associated with increased emotional positivity. The specific pattern of emotion responses appears to coincide with voice denumeration results, suggesting that both might depend in part on the same underlying percept: voice numerosity. While the relationship between voice denumeration and emotion ratings is correlational, their relationships to voice multiplicity appear to be causal.

We propose three possible accounts of why increased voice multiplicity might be associated with positive emotions: one based upon musical pleasures listeners might actually experience, one upon explicit, active interpretation of a symbolic musical surface, and a third upon on implicit processes involved with monitoring ones acoustic environment.

Experienced musical pleasure. Multi-voice music offers listeners certain musical pleasures that monophonic music does not, and this experienced pleasure might be misattributed to the musical stimuli. The simple perception of concordant harmony has long been described as a sensory pleasure. Additionally, there might be musical enjoyment involved with the successful parsing of a musical texture (Huron, 2001). Notably, one might expect that the population of undergraduates studying aural skills would be particularly likely to enjoy resolving multiple voices, raising the strong possibility that these results might not generalize beyond trained musicians.

Musical depiction. Music might be interpereted as explicitly representing happy or otherwise positive social situations. Parties, dances, celebrations, ceremonies all occur in group settings marked by positive affect. On the other hand, one could also imagine instances where high voice numerosity would not necessarily be associated with positive affect, depending on the context. A traveling merchant would surely prefer to hear the sound of one marauding warrior instead of one hundred. Perhaps an opposite voice multiplicity effect would be expected for loud, aggressive, or threatening timbres.

Environment monitoring. Finally, one could conjecture that voice numerosity is implicitly processed alongside other auditory information used continuously to monitor our auditory environment. For example, gregarious animals who live in groups often use auditory signals known as contact calls (or cohesion calls) to communicate their presence to conspecifics. One could speculate that music is sometimes processed as though it were composed of human contact calls. In this view, sounds resembling several nonthreatening human voices ought to be perceived positively, as they indicate a safe environment. However, it is unclear the degree to which this might influence listeners who are deliberately attending to the music.

NUMEROSITY CUES

If greater voice numerosity leads to more positive emotion perception, then one would expect thicker musical textures to produce this effect only to the extent listeners perceive more voices. What aspects of a musical texture might contribute to numerosity percepts?

One perceptual candidate is the density with which musical events occur, or the music's *onset density.* Simply put, music that has more frequent sounds might produce higher voice numerosity. However, not all musical events are equivalent; a note, an interval, and a three-note chord can each be sounded with a single onset, but are perceptually quite distinct. To accomodate differences in the "harmonic depth," *note density* can be defined as the number of individual notes occurring in a period of time. Because one can never have more onsets than notes, onset density can never exceed note density.

Note density and onset density for the 36 excerpts used in the second experiment are shown in Figure 9, with the shaded region indicating the degree of a typical event's harmonic depth. Both appear to be positively correlated with the voice denumeration and emotion ratings depicted in Figure 8, suggesting either note density or onset density could play a role in perceived emotional positivity. An experimental paradigm in which note density, onset density, and harmonic depth are independently manipulated could help determine their

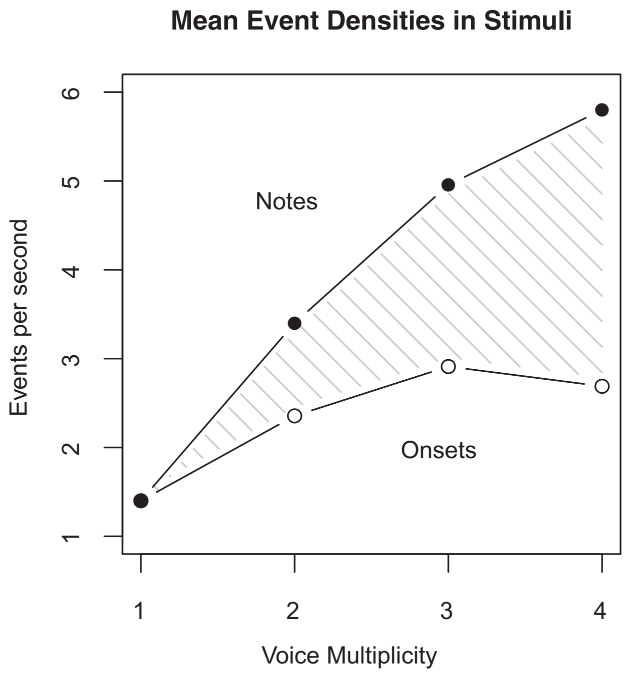

FIGURE 9. Average note and onset density in the minor-mode fugal stimuli used in the first study, expressed in events per second. Note density increases with voice multiplicity in a fairly linear relationship. Onset density, on the other hand, changes most between one-voice and two-voice textures, before leveling off. The shaded region represents the music's typical harmonic depth. Lower onset density is associated with homophonic textures, higher onset density with polyphonic textures.

relative influence.[10] Moreover, future research might fruitfully consider what specific acoustic correlates of increasing voice multiplicity (such as loudness or roughness) serve as the most salient numerosity cues.

POLYPHONICITY, HOMOPHONICITY, AND EMOTION

It is typically easier to resolve individual voices in polyphonic textures than in homophonic textures due to staggered voice entries (Huron, 1993, 2001; Rasch, 1981). Moreover, polyphonic textures tend to exhibit higher onset density in general, producing a more 'active' musical surface. Both might increase perceptual voice numerosity, and therefore lead to the perception of more positive emotions. If so, then the relative polyphonicity of stimuli might account for why Gregory et al. (1996) and McCulloch (1999) also observed a positive correspondence between textural density and emotional valence, whereas Kastner and Crowder (1990) and Webster and Weir (2005) did not. Although their descriptions are lacking, the accompaniments used by Gregory et al. (1996) and McCulloch (1999) seem to have had a polyphonic character similar to the present stimuli, leading to increased voice numerosity and more positive emotional valence perception. Meanwhile, the more homophonic music used by Kastner and Crowder (1990) and Webster and Weir (2005) would not greatly increase the music's valence over monophonic versions.

The present result that monophonic stimuli tend to evoke stronger ratings than thicker textures could help explain why Webster and Weir (2005) observed decreased positivity with increased textural density. Instead of using separate scales for sadness and happiness, these researchers asked listeners to use a single continuous valence scale from 'sad' to 'happy' to make their ratings. Their excerpts were typically perceived on the positive side of this scale, with unharmonized melodies being perceived as more positive. While these researchers took this to indicate that the harmonized melodies were perceived 'more negatively,' an alternative explanation consistent with the present results would be that the emotion ratings were simply more *neutral* (i.e., less positive).

FUTURE DIRECTIONS

The observed effect of musical texture on voice denumeration and emotion perception would benefit from generalization to other musical styles, modes, composers, and instruments. One would especially want to rigorously test the idea that perceptual voice numerosity *per se* mediates the emotion effect. Direct means of manipulating and measuring the number of musical streams listeners are able to hear could be developed, potentially exploiting pseudopolyphony or interactions with the visual modality.

The idea that homophony provides fewer numerosity cues than polyphony could be tested using specially constructed stimuli that independently manipulate note density, onset density, or onset synchrony. Specifically, the idea the onset density serves as a voice numerosity cue would be straightforward to verify experimentally.

The 'musical pleasure' explanation might be tested against the 'depiction' and 'environment monitoring' conjectures by attempting to replicate the present results using nonmusical human speech sounds. One could also test whether spoken voice denumerability exhibits similar limitations to those for musical voices, or shares the proposed numerosity cue of onset density. Additionally, examining the role of attention in a similar

---

[10] One could also note that greater rhythmic complexity and faster tempi, features related to onset density, are associated with positive valence as well (see Gabrielsson & Lindström, 2010).

paradigm could help determine the degree to which explicit interpretation of musical depiction plays a role.

Finally, task demands and implicit effects appear to play important roles in musical emotion rating paradigms. Much can be gained by carrying out several replications for any musical emotion study, each implementing a slightly modified protocol.

Author Note

Correspondence concerning this article should be addressed to Yuri Broze, School of Music, 1866 College Road, Ohio State University, Columbus, OH 43210. E-mail: broze.3@osu.edu

References

Albrecht, J. D. (2012). *Affective analysis of music using the progressive exposure method: The influence of bottom-up features on perceived musical affect* (Unpublished doctoral dissertation). Ohio State University, Columbus, OH.

Bach, J. S. (1722). *The Well-Tempered Clavier, Book 1.*

Berry, W. (1976). *Structural functions in music*. Englewood Cliffs, NJ: Prentice-Hall.

Blacking, J. (1973). *How musical is man?* Seattle, WA: University of Washington Press.

Bornstein, R. F. (1989). Exposure and affect: overview and meta-analysis of research, 1968–1987. *Psychological Bulletin*, *106*, 265–289.

Bregman, A. (1978). Auditory streaming is cumulative. *Journal of Experimental Psychology: Human Perception and Performance*, *4*, 380–387.

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

Bregman, A., & Campbell, J. (1971). Primary auditory stream segregation and perception of order in rapid sequences of tones. *Journal of Experimental Psychology*, *89*, 244–249.

Burnett, S., Bird, G., Moll, J., Frith, C., & Blakemore, S.-J. (2009). Development during adolescence of the neural processing of social emotion. *Journal of Cognitive Neuroscience*, *21*, 1736–1750.

Cacioppo, J. T., & Patrick, W. (2008). *Loneliness: Human nature and the need for social connection*. New York: W. W. Norton.

Cambouropoulos, E. (2008). Voice and stream: Perceptual and computational modeling of voice separation. *Music Perception*, *26*, 75–94.

Cheatham, P. G., & White, C. T. (1954). Temporal numerosity: Iii. auditory perception of number. *Journal of Experimental Psychology*, *47*, 425–428.

Clendinning, J. P., & Marvin, E. W. (2005). *The musician's guide to theory and analysis*. New York: W. W. Norton and Company.

Cooper, P. (1981). *Perspectives in music theory: An historical-analytical approach*. New York: Harper & Row.

Davies, M. (2008–). *The corpus of contemporary American English: 450 million words, 1990-present*. Available from http://corpus.byu.edu/coca/ (Accessed May 2012)

Davis, S. (2006). Implied polyphony in the solo string works of J. S. Bach: A case for the perceptual relevance of structural expression. *Music Perception*, *23*, 423–446.

Davison, L. L., & Banks, W. P. (2003). Selective attention in two-part counterpoint. *Music Perception*, *21*, 3–20.

Dehaene, S. (1997). *The number sense: How the mind creates mathematics*. New York: Oxford University Press.

Descoeudres, A. (1921). *Le développement de l'enfant de deux à sept ans* [Childhood development from age two to seven]. Paris: Delachaux & Niestle.

Eisenberg, N. (2000). Emotion, regulation, and moral development. *Annual Review of Psychology*, *51*, 665–697.

Ekman, P. (1973). Darwin and cross-cultural studies of facial expression. In P. Ekman (Ed.), *Darwin and facial expression* (pp. 1–83). New York: Academic Press.

Fujioka, T., Trainor, L. J., Ross, B., Kakigi, R., & Pantev, C. (2005). Automatic encoding of polyphonic melodies in musicians and nonmusicians. *Journal of Cognitive Neuroscience*, *17*, 1578–1592.

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

Gabrielsson, A., & Lindström, E. (2010). The role of structure in the musical expression of emotions. In P. N. Juslin & J. A. Sloboda (Eds.), *Handbook of music and emotion: Theory, research, applications* (pp. 367–400). New York: Oxford University Press.

Gregory, A., Worral, L., & Sarge, A. (1996). The development of emotional responses to music in young children. *Motivation and Emotion*, *20*, 341–349.

von Helmholtz, H. (1863). *Die Lehre von den Tonempfindungen als physiologische Grundlage für die Theorie der Musik* [On the sensations of tone as a physiological basis for the theory of music]. Braunschweig: F. Vieweg.

ten Hoopen, G., & Vos, J. (1979). Effect on numerosity judgment of grouping of tones by auditory channels. *Perception and Psychophysics*, *26*, 374–380.

Huron, D. (1989a). Characterizing musical textures. In T. Wells & D. Butler (Eds.), *Proceedings of the 1989 international computer music conference* (pp. 131–134). San Francisco, CA: Computer Music Association.

Huron, D. (1989b). Voice denumerability in polyphonic music of homogeneous timbres. *Music Perception*, *6*, 361–382.

Huron, D. (1993). Note onset asynchrony in J. S. Bach's two-part inventions. *Music Perception*, *10*, 435–444.

Huron, D. (2001). Tone and voice: A derivation of the rules of voice-leading from perceptual principles. *Music Perception*, *19*, 1–64.

Ilie, G., & Thompson, W. F. (2006). A comparison of acoustic cues in music and speech for three dimensions of affect. *Music Perception*, *23*, 319–329.

Jevons, W. (1871). The power of numerical discrimination. *Nature*, *3*, 363–372.

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

Kagan, J. (1981). *The second year: The emergence of self-awareness*. Cambridge, MA: Harvard University Press.

Kastner, M. P., & Crowder, R. G. (1990). Perception of the major/minor distinction: IV. emotional connotations in young children. *Music Perception*, *8*, 189–202.

Kaufman, E., Lord, M., Reese, T., & Volkmann, J. (1949). The discrimination of visual number. *American Journal of Psychology*, *62*, 498–525.

Kivy, P. (1980). *The corded shell.* Princeton, NJ: Princeton University Press.

Laitz, S. G. (2012). *The complete musician: An integrated approach to tonal theory, analysis, and listening* (3rd ed.). New York: Oxford University Press.

Lakoff, G. (1987). *Women, fire and dangerous things: What categories reveal about the mind.* Chicago, IL: University of Chicago Press.

Lewis, M., & Michalson, L. (1983). *Children's emotions and moods: Developmental theory and measurement*. New York: Plenum.

Lewis, M., Sullivan, M. W., Stanger, C., & Weiss, M. (1989). Self development and self-conscious emotions. *Child Development*, *60*, 146–156.

McCulloch, R. (1999). *Modality and children's affective responses to music* [Undergraduate project for Perception and Performance course (Ian Cross, instructor)]. Cambridge. Available from http://www.mus.cam.ac.uk/ic108/PandP/McCulloch99/McCulloch99.html

Merriam, A. P. (1964). *The anthropology of music*. Evanston, IL: Northwestern University Press.

Minzenberg, M. J., Poole, J. H., & Vinogradov, S. (2006). Social-emotion recognition in borderline personality disorder. *Comparative Psychiatry*, *47*, 468–474.

Nettl, B. (2005). *The study of ethnomusicology: Thirty-one issues and concepts*. Urbana, IL: University of Illinois Press.

Ollen, J. E. (2006). *A criterion-related validity test of selected indicators of musical sophistication using expert ratings* (Unpublished doctoral dissertation). Ohio State University.

Parncutt, R. (1993). Pitch properties of chords of octave-spaced tones. *Contemporary Music Review*, *9*, 35–50.

Rasch, R. A. (1981). *Aspects of the perception and performance of polyphonic music*. (Unpublished doctoral dissertation). Elinkwijk, Utrecht, The Netherlands.

Rosch, E. (1973). Natural categories. *Cognitive Psychology*, *4*, 328–350.

Rosch, E. (1975). Cognitive representation of semantic categories. *Journal of Experimental Psychology*, *104*, 192–233.

Rousseau, J. J. (1782). Essai sur l'origine des langues. In *Collection complete des oeuvres de J. J. Rousseau, citoyen de Genève* [The complete works of J. J. Rousseau] (Vol. 8, pp. 355–434). Geneva.

Rowell, L. (1983). *Thinking about music: An introduction to the philosophy of music*. Amherst, MA: University of Massachusetts Press.

Shaver, P., Schwartz, J., Kirson, D., & O'Connor, C. (1987). Emotion knowledge: Further exploration of a prototype approach. *Journal of Personality and Social Psychology*, *52*, 1061–1086.

Smith Brindle, R. (1966). *Serial composition*. Oxford: Oxford University Press.

Spencer, H. (1857). The origin and function of music. *Fraser's Magazine*, *56*, 396–408.

Stipek, D. J., Gralinski, J. H., & Kopp, C. B. (1990). Self-concept development in the toddler years. *Developmental Psychology*, *26*, 972–977.

Tangney, J. P., Stuewig, J., & Mashek, D. J. (2007). Moral emotions and moral behavior. *Annual Review of Psychology*, *58*, 345–372.

Teroni, F., & Deonna, J. A. (2008). Differentiating shame from guilt. *Conscious Cogition*, *17*, 725–740.

Thompson, W. F., Marin, M. M., & Stewart, L. (2012). Reduced sensitivity to emotional prosody in congenital amusia rekindles the musical protolanguage hypothesis. *Proceedings of the National Academy of Sciences*, *109*, 19027–19032.

Tracy, J. L., & Robins, R. W. (2007). The nature of pride. In J. L. Tracy, R. W. Robins, & J. P. Tangney (Eds.), *The self-conscious emotions: Theory and research*. New York: Guilford.

Tracy, J. L., Robins, R. W., & Tangney, J. P. (Eds.). (2007). *The self-conscious emotions: Theory and research*. New York: Guilford.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*, 1124–1130.

Verlet, B. (1993). *Well-Tempered Clavier, Book I* [Compact disc recording]. Astré-Auvidis.

Webster, G. D., & Weir, C. G. (2005). Emotional responses to music: Interactive effects of mode, texture, and tempo. *Motivation and Emotion*, *29*, 19–39.

Weiss, R. S. (1975). *Loneliness: The experience of emotional and social isolation*. Boston, MA: MIT Press.

White, C. T. (1963). Temporal numerosity and the psychological unit of duration. *Psychological Monographs: General and Applied*, *77*, 1–37.

Williams, L. A., & DeSteno, D. (2008). Pride and perseverance: The motivational role of pride. *Journal of Personality and Social Psychology*, *94*, 1007–1017.

Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology*, *9*, 1–27.